Systematic Literature Review of Patient Data Privacy and Security for Software System Development

A Thesis Presented to

Department of Computer Science & Software Engineering
Faculty of Basic & Applied Sciences

In Partial Fulfillment of the requirement for the degree

Of

Master of Sciences (Software Engineering)

By

Isma Masood

International Islamic University Islamabad

1. Software compatibility

2. Software reliability

بسم الله الرحمن الرحيم

In the Name of Allah, the Most Gracious, the Most Merciful

(2012)

International Islamic University, Islamabad

Faculty of Basic & Applied Sciences

Department of Computer Science & Software Engineering

Dated:

FINAL APPROVAL

It is certified that we have read the thesis, entitled **"Systematic Literature Review of Patient Data Privacy and Security for Software System Development"**, and submitted by Isma Masood Reg. No. 196-FBAS/MSSE/S08. It is our judgment that this thesis is of sufficient standard to warrant its acceptance by the International Islamic University Islamabad for MS Degree in Software Engineering.

PROJECT EVALUATION COMMITTEE

External Examiner:

Dr. Arshad Ali Shahid

HOD, Department of Computer Science

FAST, nuces.

Internal Examiner

Salma Imtiaz

Assistant Professor, Department of Computer Science,

Faculty of Basic and Applied Sciences,

International Islamic University Islamabad, Pakistan

Supervisor:

Dr. Saad Zafar

Dean, Faculty of Computing,

Riphah International University

Islamabad, Pakistan.

Co-supervisor

Mr. Imran Saeed

Assistant Professor, Department of Computer Science,

Faculty of Basic and Applied Sciences,

International Islamic University Islamabad, Pakistan

Acknowledgement:

I am very thankful to ALLAH who blessed me in the course of my life and especially during this thesis. It was only with ALLAH's help that I am able to complete my work.

I would like to thank Dr. Saad Zafar for their continued support, guidance and patience for the completion of this work.

I express my gratitude to my lovely mother, friends, and peers for their extraordinary support during my research work. I am thankful to Mr. Imran Saeed, Mrs Bushra Malik, Mrs Noureen, Mr. Muhammad Saud and Mrs Summera Saleem for their help in my thesis.

I am also very grateful to Professor Barbara Kitchnham, Dr. Mehmood Niazi and Dr. Naveed Ikram for their expert opinions.

Declaration:

I hereby declare and affirm that this thesis neither as a whole no as part thereof has been copied out from any source. It is further declared that I have completed this thesis entirely on the basis of my personal effort, made under the sincere guidance of my supervisors. If any part of this report is proven to be copied our or found to be a reproduction of some other, we shall stand by the consequences. No portion of the work presented in this report has been submitted in support of an application for other degree or qualification of this or any other University or Institute of learning.

Isma Masood

196-FBAS/MSSE/S08

Dedication:

I dedicate my work to:

My Parents
&
My Teachers

Abstract:

Objective: The purpose of this systematic review is to identify and classify patient data privacy and security solutions proposed for Health Information Systems (HIS) according to predefined criteria.

Materials and Methods: An automated search was performed using four well-known electronic databases to find studies explicitly addressing any of the patient data privacy or security concerned published during the period ranging from 2000 to 2011.

Results: From 633 citation references, 123 full text papers were read and 77 independent studies were selected on the basis of our inclusion, exclusion and quality criteria. These studies focused primarily on issues related to software architecture and design (64%), software implementation (27%), and requirements lifecycle (9%). No study was found related to software testing and maintenance.

Discussion: The studies were mapped on predefined privacy principles. A single study was often found to be covering multiple privacy principles, with Security Safeguard and Controls (76/77) getting the most attention, followed by Individual Participation and Control (18/77). Most studies reported solutions for Decentralized HIS architecture (29/77) as compared to a Centralized HIS (7/77).

Conclusion: A growing trend was observed in the literature for proposing solutions for patient data privacy in software-based systems. The main focus of the studies was on Software Architecture and Design, whereas, considerably less attention was paid to the Requirements Engineering phase. All of the selected studies did cover the complete spectrum of predefined privacy principles in one way or the other, with the Security Safeguard and Control principle getting the most coverage.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

EHRs	Electronic Health Records
HIS	Health Information System
SDLC	Software Development Lifecycle
SLR	Systematic Literature Review
RE	Requirement Engineering
SA&D	Software Architecture and Design
IMP	Implementation

1 INTRODUCTION

This chapter includes brief introduction of this thesis, motivation and proposed work.

1.1 Introduction

Last three decades show tremendous increase in collection and use of health information. Environmental forces with advancement in technology have played vital role which leads an unparalleled growth in depth and breadth of the collection and use of the health information. This collected health information includes diagnose and testing information with person's family history, history of diseases and treatments, genetic testing, sexual orientation and practices, history of drugs used, and testing for sexually transmitted diseases [1]. The amount of primary and secondary uses of health information has grown terrifically. The primary use of health information included patient care delivery, support and management of billing with reimbursement [1]. Whereas, secondary uses improving effectiveness within healthcare system, conduct medical research, drive public policy development and administration, to justify payment of services rendered to payer organizations, for consumer to make conversant choices about health plans and health providers, to provide effective health services which can be tracked and evaluated [1, 2]. Figure 1 shows complete flow of health information.

1.1.1 Electronic Health Records (EHRs)

Literature reported large amount of terms to refer electronically processed health information like Personal Health Information (PHI), Electronic Medical Records (EMRs), Health Records (HRs) and Electronic Health Records (EHRs) etc. However, the term Electronic Health Records (EHRs) used to refer health information in this review.

The institute of Medicine (IOM) published a landmark report [3] in 1991. This report refers EHRs as computer-based patient record (CPR). Reports define CPR as: "electronic patient records that reside in a system designed to support users through availability of complete and accurate data, practitioner reminders and alerts, clinical decision support systems, links to bodies of medical knowledge, and other aid" [3]. Later Richard Dick one of the editors of same report defines CPR as" a representation of all of a patient's data that one would find in the paper-based record, but in a coded and structured, machined-readable form" [4]. These definitions clearly defined the purpose, nature and use of EHRs.

1.1.2 Health Information System (HIS)

The use of computer systems in clinical data management activities has been reported in literature since late 1950s. The most common systems involve in these activities are master patient index, pharmacy information system, radiology information system, picture archiving system, nursing information system, health information system (HIS), chart management/medical records systems, practice management system and laboratory information system. Health information system (HIS) used in healthcare sectors are responsible to collect, store, process and update the health information [5]. A report published by Institute of Medicine in 2003 reported "Key Capabilities of EHR system" [5]. A system includes: (1) longitudinal collection of health information about persons (patients and healthcare provider); (2) immediate access to electronic health information by authorized users; (3) knowledge and decision support to enhance quality, safety and efficiency of patient care; (4) support for efficient processes for health care delivery.

This definition of the system encompasses all of the concepts and functionality proposed originally for the EHRs.

1.1.3 Patient Data Privacy and Security

Privacy is considered primary relationship between patient and physician [1]. This relationship is directly influenced by quality and amount of information provided by patients. The quality and quantity of EHRs unswervingly leads individual's life and financial well begin [2]. Frequently, the terms privacy and security intermingle, resulting in misuse. Therefore, context of this research defined privacy as "The right of individual to limit the access of their EHR by determining at what time, in what way, and to what extent" [6]. Similarly, security as "Protection measures that safeguards data, computer program and communication channel from undesired occurrences and exposures" [7]. Electronically shared EHRs within healthcare sector and other organizations (as shown in Figure 1) are receiving threats to patient data privacy and security, in a survey [1] these threats are categorized as organizational and systematic threats. In another study [8] information security threats in network system of medical organizations are categorized in according to events i.e.: natural events, external events and internal events. Threats to patient data privacy and security become a major cause of inaccuracies and improper disclosure of information which threaten individual's personal life and financial well being [2]. Therefore, to protect patient data privacy and security many countries have implemented different laws and policies to secure EHRs [9]. In addition, a number of standards, policies and regulations have also emerged to govern the use of EHRs in a manner that patient data privacy and security is not violated. Any software system that is developed today must adhere to the new emerging requirements. To bridge this gap between different patient privacy rules, regulations and policies, Markle Foundation [10] has proposed a Common Framework for uniform implementation of health information exchange across the health sector. Markle Foundation [10] works for the advancement of health and national security through information and information technology in the United States of America. One of the major objectives of the Common Framework is to ensure patient privacy with security and seamless connectivity among various organizations related to the health sector. The framework is based on a set of 9 principles derived from laws and policies already implemented in Canada, USA and UK [9]. These nine principles are outline below: *Openness and Transparency:* "There should be a general policy of openness about developments, practices, and policies with respect to personal data. Individuals should be able to know what information exists about them, the purpose of its use, who can access and use it, and where it resides".

Purpose Specification and Minimization: "The purposes for which personal data are collected should be specified at the time of collection, and the subsequent use should be limited to those purposes or others that are specified on each occasion of change of purpose".

Collection Limitation: "Personal health information should only be collected for specified purposes, should be obtained by lawful and fair means and, where possible, with the knowledge or consent of the data subject".

Use Limitation: "Personal data should not be disclosed, made available or otherwise used for purposes other than those specified".

Individual participation and control: "Individuals should control access to their personal information. Individuals should be able to obtain from each entity that controls personal health data, information about whether or not the entity has data relating to them. They should have the right to:

- Have personal data relating to them communicated within a reasonable time (at an affordable charge, if any), and in a form that is readily understandable
- Be given reasons if a request (as described above) is denied, and to be able to challenge such denial; and
- Challenge data relating to them and have it rectified, completed, or amended".

Data Integrity and Quality: "All personal data collected should be relevant to the purposes for which they are to be used and should be accurate, complete, and current".

Security Safeguards and Control: Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification, or disclosure.

Accountability and Oversight: "Entities in control of personal health data must be held accountable for implementing these information practices".

Remedies: "Legal and financial remedies must exist to address any security breaches or privacy violations".

1.1.4 Systematic Literature Review (SLR)

According to guidelines of Kitchenham a (SLR), is defined as:

"A form of secondary study that uses a well-defined methodology to identify, analyze and interpret all available evidence related to a specific research question in a way that is unbiased and (to a degree) repeatable" [11]. Researchers are able to make a meticulous review on their topic/questions by SLR [12]. SLR is carried out using a pre-defined search strategy [11]. A predefined search strategy makes this literature review a systematic and a fair one [12]. Therefore, its meticulous and fair nature makes this literature review unique from an ordinary literature review. SLR is also considered one of the best way to support the Education and Research [13]. Due to its systematic and fair review for literature SLRs can also help in the preparation of text books and course materials [14] . The detail description of SLR methodology is given in chapter 3. Literature reported a number of SLRs (for e.g. [13, 15-19]) on other topics of software engineering. However, to best of our knowledge, we could not find SLR on topic of "Patient Data Privacy and Security for Software System Development".

1.2 Motivation and Proposed Work

One of the challenges which are facing healthcare sector for designing HIS, right and proper privacy and security of EHRs [1]. Due to this challenge, a lot of work (for e.g. [20-26]) has been carried out to advance the state of patient data privacy and security. Researchers are proposing solutions for development of HIS in order to secure and maintain privacy of EHRs. However, focus of these solutions is software architecture and design phase of software development lifecycle (SDLC) as compare to other phases of SDLC. This published literature also shows that research community who are improving privacy and security of EHRs in different perspective e.g. exchange of EHRs in cross domain via internet and EHRs security in biomedical data analysis etc.

On the other hand, Health Information Systems (HIS) developed for healthcare sector counter some principles of privacy and security laws and policies. Furthermore, HIS followed these principles unable to incorporate all contents of these principles. Consequently, coverage of these privacy principles still need for effective integration in HIS. As a rule, before initializing a new research activity, researchers usually conduct an informal literature review on their topic of interest. But, there might be a chance that solutions to the research questions of their interest already exists in literature or another chance of not finding the appropriate solutions to answer their research questions as the solutions are scattered on internet. Due to this scattered nature of literature on internet, there is a chance of reinventing the wheel. So, there is a need to perform formal literature review before starting any new research activity. After this formal literature review, researchers are able to find their idea in literature and move further before performing any new research activity.

To answer the above stated problems, there is a need to systematically collect, analyze and synthesize the evidences related to patient data privacy and security for software system development that are present in literature.

1.3 Research Questions

The goal of this study is to conduct a systematic review which retrieves practical solutions proposed for development of HIS and mapped these solutions against aforementioned privacy and security principles [10]. The outcome of this study may be beneficial for the software developer, researchers and vendors who are involved in developing effective HIS systems. Specifically, in this study we aim to answer the following two research questions:

1) Which solutions of patient data privacy and security have been proposed for software system development?
2) Can we categorize these solutions against the privacy and security principles proposed by Markel Foundation [10]?

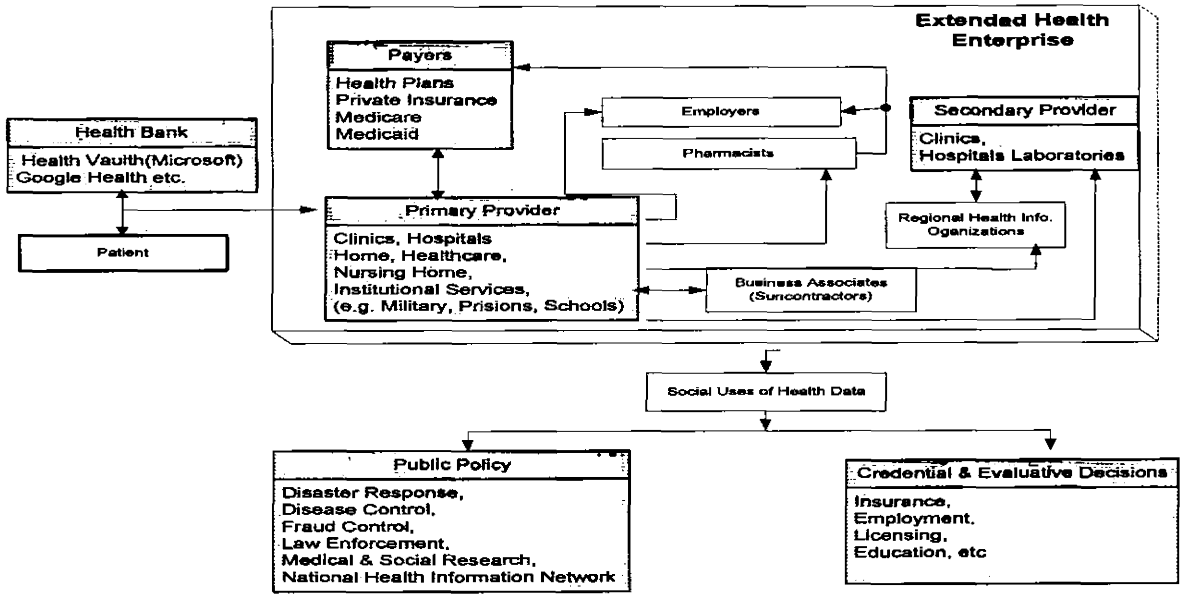

Figure 1 Graphical view for flow of EHRs adopted from [1]

1.4 Organization

This chapter gives the brief introduction to the topic patient data privacy and security. Chapter 2 provides background and related work. Chapter 3 describes the research methodology, research questions and steps of review process in details. Our 4th chapter includes the results and analysis. In chapter 5 identification and analyzes of validity threats are presented. Finally, the conclusion accompanied with discussions and future work has been conveyed in Chapter 6.

2 Background

This chapter describes background including problems of patient data privacy and security, recent advancements to secure privacy and security of patient's data, overview of systematic literature review (SLR) and related work, for better understanding of reader.

2.1 Treats to Patient Data Privacy and Security

As mentioned before, patient data privacy and security receiving threats [1]. These threats become major cause of inaccuracies in EHRs which directly influence patient health and financial well begin [1]. However, literature reported some important and serious threats to privacy and security of EHRs. We have categorized each threat according to their types.

2.1.1 Treats to Information Privacy

The following are the common threats to information privacy.

- **Insider accidental disclosures/errors**
  This is the probably the most common threat to informational privacy. In this case the employer, due to lack of knowledge of organization policy or some other reason, releases private information to unauthorized individuals [22, 27].
- **Abuse by insiders of their access privileges**
  In this case individuals who have authorized access to target data violate the trust associated with that access. For example, an employee who reviews the medical information of a colleague, family member, or friend for non-healthcare delivery purposes [28].
- **Insider unauthorized access**
  This includes those accesses where an employee may have access to the information system but targets access to unauthorized information through exploitation of system vulnerability or other means. Frequently this type of action will be for spite or profit [2].
- **Outside intruders**
  Individuals who do not have authorized access to the system but gain such access through exploitation of system vulnerabilities or other means to explore information stores or mount attacks to damage systems or disrupt operation [1].

2.1.2 Treats to Information Integrity

Several of the treats to informational privacy also apply to those related to information integrity.

- **Insider accidental error**

  In this case the employee inputs incorrect values or collects data inappropriately. This may be due to lack of knowledge or to poor system constraints [29].

- **Insider Malicious attack**

  In this case the employer purposefully corrupts data for which there may be unauthorized access [28].

- **Intruder attack**

  In this case an intruder gains access to the system by exploitation of system vulnerability or other means and either accidently or purposefully corrupts or destroys data [20].

- **Software failure**

  Failure of application or system software due to performance, inadequate code, or other reason fails to adequately protect data integrity [30].

- **Strategic attack**

  Includes purposeful attack, such launching of a computer virus that corrupts or destroys data [6].

2.1.3 Treats to Information Reliability

Threats to information reliability involve a wide spectrum of incidents ranging from natural disasters to human error. Among these are:

- **Natural hazards**

  This includes such things as earthquakes, tornadoes, ice storm, fires, floods, electrical storms, and so on [6].

- **Equipment and software failures**

  This includes hardware breaks downs and software failures that cause unexpected system suspension or shutdown [6].

- **Human error**

  This includes any human error that would cause hardware or software to improperly function, causing unexpected system distribution or shutdown [6].

- **Theft, malice or strategic attack**

  Includes purposeful theft or attack on any component of the information system with the intent of accusing system disruption or shutdown [6].

2.2 Recent Advancements for Privacy and Security of Patient Data

Following are some important and recent advancement secure privacy and security of patient's data.

- An architecture proposed [21] for accurate and reliable sharing of EHRs by satisfying patient privacy needs. This architecture utilizes pseudonym identifiers to empower patients at their EHRs [21].
- A unified framework combines techniques of binning and digital water marking to achieve the dual goals i) patient data privacy and ii) copyright protection for secondary use [31].
- Dalley A et al [29] model allow patients to withdraw and award access rights to GP by using I-Keys hardware. As a result the EHRs can be access at anytime and at any location on patient's choice.
- ChiperMe technology based architecture of Hensen I [32] store EHRs in separate boxes. The id of separate EHRs deposit boxes are in control of patients therefore, patients are fully empowered at their records.
- Cheng T-L [33] proposed a key management scheme to make mobile agents more secure and efficient access control.
- An access control model based framework designed for sharing of patient profiles in distributed healthcare environment. The proposed model captures the dynamic behavior and customizes access control according to access rights [34].

- In another study, digital rights management is used to secure the transmission of EHRs. The proposed solution disclosed patient data on need-to-know principle as defined by patients [35].
- Similarly, a flexible cryptographic key management solution is proposed to comply with the HIPAA regulations in order to protect privacy and security of EHRs [36].
- Deng M et al [37] proposed context-specific scheme for 'Federated Electronic Healthcare'(FEH). This scheme proposed identity management framework based on cryptographic algorithm. Hence, identities are issues according to the specific context for controlled authorization. The existing scheme in FEH is improved by introducing algorithm. This improvement do not privileged the attacker to correlate EHRs on large scale.
- An e-prescription system [38] is proposed to protect patient privacy in drug prescription buy utilizing smart cards. The identity of patients and doctors are unlinkable by pseudonymity. This unlink-ability hides patient identification.

2.3 Systematic Literature Review

“As a research area matures there is often a sharp increase in the number of reports and results made available, and it becomes important to summarize and provide overview” [39]. According to Kitchenham, “Systematic Literature Review (SLR) is a means of identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area, or phenomenon of interest" [11].

During SLR method primary studies are identified, published in area of interest. These primary studies become source for review under related research questions. Kitchenham [11] have adopted the concept of performing SLRs in software engineering from medical sciences. SLR is a form of secondary study which based on published studies of recently gained attention area by researchers in software engineering related areas [39-41].

According to Petersen [39] under SLR method primary studies of related research questions have reviewed for in depth analysis of results, methodologies and context. There are many reasons for conducting SLRs. However, most common mention by Kitchenham [11] are as follows.

- "To summarize the existing evidence concerning a treatment or technology e.g. to summarize the empirical evidence of the benefits and limitations of a specific agile method.
- To identify any gaps in current research in order to suggest areas for further investigation.
- To provide a framework/background in order to appropriately position new research activities."

As systematic Literature Review (SLR) is a predefined research methodology, covered complete literature in unbiased and auditable manner. Therefore, SLR allow general conclusion based on extensive range of situations and contexts [11, 39]. The main rational for performing SLR is lack of scientific value in literature review performed by different researchers [11]. However, SLR is a laborious job which required lot of efforts [11, 39].

Following are some important features which differentiate Systematic Literature Review from conventional literature review.

- ✓ A predefined protocol is designed to perform SLR which specifies research questions and methods used for conducting SLR.
- ✓ The search strategy of SLR allows documenting all relevant literature, resulting completeness and reliability.
- ✓ Inclusion/exclusion criteria are designed for research questions which allow careful selection of relevant primary studies.
- ✓ The quality of extracted data is assessed against predefined criteria [11].

2.4 Related Work

As, we did not find any SLR on patient data privacy and security so, our related work is divided into three sections. The first section will discuss some work related to patient data privacy and security only. The second group will discuss some work related to systematic literature review in healthcare domain however; our third group reports some SLRs in software engineering field.

2.4.1 Patient Data Privacy and Security

In 2010 a survey [1] published which critically survey the literature on information security and privacy in healthcare by covering various research methodologies such as Design Research

(Algorithm, Architecture/Framework, Measurement, Experiment/Simulation, Conceptual Modeling, Prototyping), Qualitative Research (Policy Report, Topical Discussion/Analysis, Case Study, Theory Building, Interview) and Quantitative Research (Empirical Study with primary data, Empirical study with secondary data, Economic Modeling, Mathematical /Statistical Modeling). For categorization of papers reviewed in this survey they conducted a multidisciplinary search in a diverse set of publications and fields including information systems, health informatics, public health, medicine, law and articles in popular trade publications and reports. On the basis of these research methodologies four primary research domains in the healthcare information security and privacy are defined that intersect with corresponding four domains of information-systems-related research in healthcare. This survey only highlights threats to information privacy and security in domains like privacy concerns among healthcare consumers, providers' perspectives of regulatory compliance, information access control, data interoperability and information security, Information security issues of e-Health, information security for authorized data disclosure, information integrity in healthcare and adverse effects, financial risk and fraud control, regulatory implication to healthcare practice and information securl ity risk management. This survey shows that there is an emerging trend in information security and privacy in healthcare sector. Therefore, there is a great need of literature synthesis for privacy and security of EHRs in order to integrate proper requirements in HIS.

2.4.2 Systematic Literature Reviews (SLRs) in Healthcare Domain

Literature has showed huge number of SLRs in other areas of Healthcare Domains. Poisannt [42] performed an SLR in order to examine the impact of EHRs on documentation time of nurses and physicians. In this SLR total 23 papers were identified on the basis of inclusion criteria. In another study Calvin K [43] identified variables involved in promoting consumer health information technology (CHIT). In total, 52 articles met selection criteria. Similarly, to combine existing evidence about adaptation of HIS 50 studies were reviewed published during 1994-2008 [44].

2.4.3 Systematic Literature Reviews (SLRs) in Software Engineering

As mentioned before, due to scattered nature of literature different areas of software engineering have been summarized and classified by using SLR methodology. Through secondary studies like systematic mapping and Systematic Reviews researchers can have quick and easy review on a specific research topic without searching all published studies in that area. The idea to perform SLRs in software engineering adopted from medical field by Kitchenham [11]. There are number of systematic reviews (e.g. [45] [46, 47]) in different areas of software engineering but not really related to the patient data privacy and security and research questions of this thesis.

Peterson et al [39] has discussed the use of systematic mapping, review and their procedure in software engineering related areas. While Michael et al [19] have reviewed 148 papers published between 1991 and 2008 to address evaluation and measurement of Software Process Improvement, Magne Jørgensen and Martin Shepperd [48] performed a SLR on Cost Estimation and reviewed 304 papers published in 76 journals. Tore Dybå et al [40] also systematically identified and analyzed all published studies on pair programming. Similarly, Tracy Hall, et al [49] reviewed published studies on motivation in software engineering between 1980 and 2006.

All of the above mentioned studies focus on a specific category of software engineering area. However, literature reported large number of solutions on software system development to maintain privacy and security of patients EHRs. No single study reported for identification and classification of published literature on patient data privacy and security. Therefore, a general systematic review is in this area will synthesized the evidence for future progress.

To cover all aspects of patient data privacy and security for software system development related research papers in software engineering, two main perspectives have been defined for this systematic review which are an authenticated predefined privacy and security principles [10] and software development Lifecycle (phases). More explanation about this has been given in chapter 3 and 4.

3 RESEARCH METHODOLOGY

In this chapter the procedure and the design of SLR of this thesis has been described. This chapter mainly highlights research questions, mapping study, selection procedure and data extraction phases of SLR.

3.1 Systematic Literature Review

Figure 2 shows detail steps and procedure used to conduct SLR for this thesis. The SLR methodology includes three major phases 1) Planning, 2) Conducting and 3) Documenting. The detail descriptions of each phase are mentioned below.

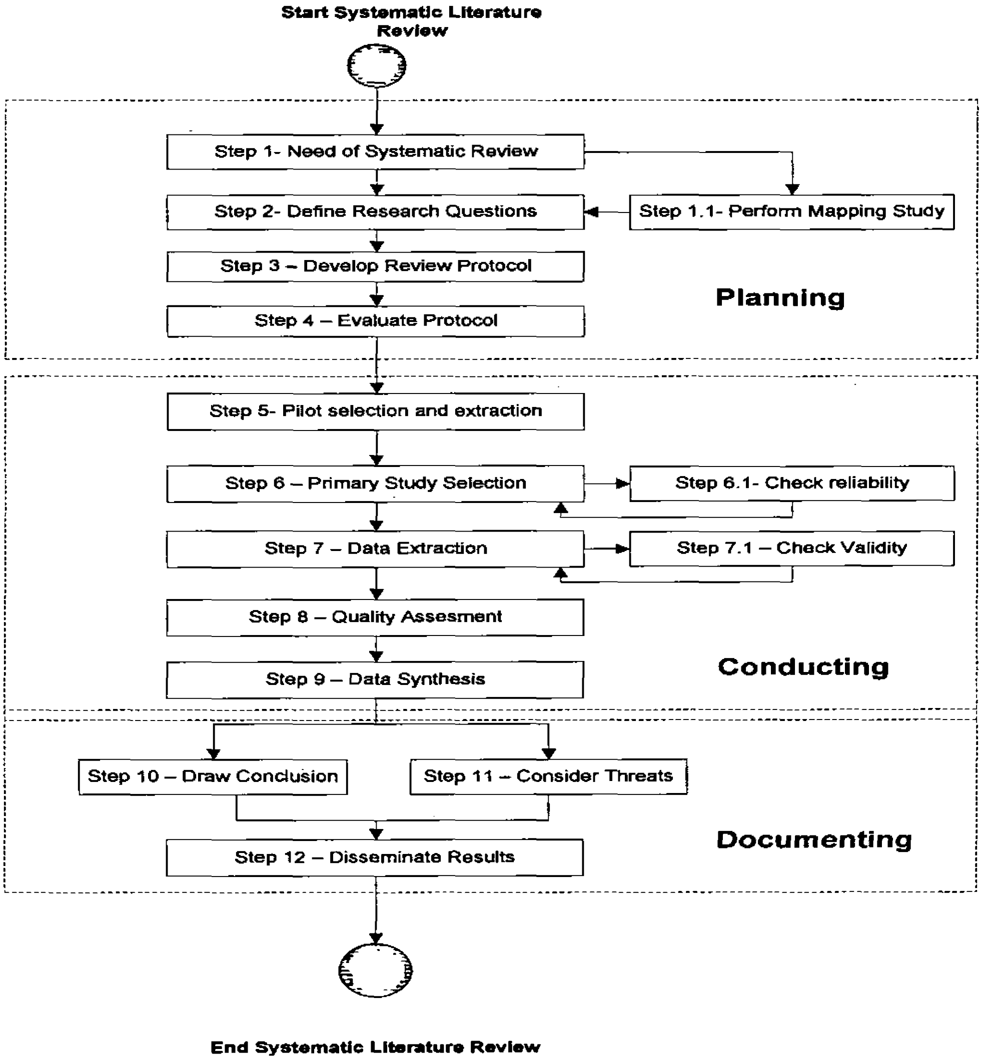
Start Systematic Literature Review
Step 1- Need of Systematic Review
Step 2- Define Research Questions
Step 1.1- Perform Mapping Study
Step 3 – Develop Review Protocol
Step 4 – Evaluate Protocol
Planning
Step 5- Pilot selection and extraction
Step 6 – Primary Study Selection
Step 6.1- Check reliability
Step 7 – Data Extraction
Step 7.1 – Check Validity
Step 8 – Quality Assesment
Step 9 – Data Synthesis
Conducting
Step 10 – Draw Conclusion
Step 11 – Consider Threats
Step 12 – Disseminate Results
Documenting
End Systematic Literature Review

Figure 2 Steps of Systematic Literature Review

3.1.1 Planning

In this study, we have conducted a systematic Literature Review (SLR) to retrieve practical solutions proposed for development of HIS systems for healthcare sector and mapped these solutions against predefined privacy and security principles [10]. The outcome of the study may be beneficial for the software developers and researchers who are involved in developing effective HIS systems.

3.1.1.1 Need for SLR

The primary reason of conducting a SLR on patient data privacy and security for software system development is that up to our best knowledge, no SLR has been published yet. Therefore, we decided to conduct SLR on this important and timely research topic. The major reasons behind this work are:

- Governments of international community's is busy in developing laws and policies to maintain privacy and security of patients EHRs [9, 10]. Whereas, research community is busy in developing solutions to protect patient privacy and security due to wide exchange of EHRs in different organizations [1]. But the problem is that these proposed solutions are not classified against principles/guidelines of any privacy law or policy. As, proposed solutions are random so, unable to follow all contents of each guideline of privacy policy or law.
- Solution on patient data privacy and security also need classification according to SDLC phases to access the progress for HIS development.
- Precisely, to help:
  - ✓ Vendors → able to get practical solutions on patient data privacy and security for software system development.
  - ✓ Researchers→ to get state of the art on patient data privacy and security which leads towards future progress by highlighting new research topics. Therefore, endless efforts will be stop for doing same kind of improvements.

3.1.1.1.1 Mapping Study

In the field of software engineering, systematic mapping study is quite a new research method adopted from other disciplines by Kitchenham [11]. The purpose to perform this mapping study was to see feasibility of systematic review for this area. Guidelines for this mapping study have been adopted from [50, 51].

We have conducted a mapping study to identify what type of contributions are available on patient data privacy and security for software system development and is this possible to categorized these solutions against selected authenticated policy principles [52] for patient data privacy and security.

The main motivation to perform mapping study was to:

- ✓ To target relevant publication channels.
- ✓ To cover breath of the focused area.
- ✓ Estimate time span for conducting the SLR.
- ✓ Identify a classification scheme to synthesize the results.
- ✓ To derive relevant terms for search string for conducting the SLR.
- ✓ To check the feasibility of research questions phrased to conduct SLR.

Definition of Research Questions: The major aim of systematic mapping studies is to provide an overview of our research area by identifying the quantity and type of research available to perform systematic literature review. To achieve this goal we check the feasibility of our two research questions:-

RQ1 which solutions of patient data privacy and security have been proposed for software system development?

RQ2 Can we categorize these solutions using the Markle Foundation's Common Framework [10]?

The systematic mapping process

For our mapping study we have followed the guidelines provided in [7, 8]. The process of mapping study has been divided into three stages. At stage 1, we have defined the scope, the search strategy and the selection criteria. Stage 2 covered selection of primary studies by applying the search strategy and the selection criteria. Finally, in Stage 3, the selected studies are classified into the different categories. The results of our mapping study have been published in international conference [53]. For detail mapping study see appendix 2.

3.1.1.2 Search String

Literature has addressed large number of studies on privacy and security of images, data, sensor networks, wireless communication and in home monitoring patients. Therefore, there was a great need to define the scope of SLR. For this purpose a detailed mapping study have performed as suggested by Kitchenham [11].

We used these results in finalizing the SLR protocol which is described next. The results of mapping study published in Oct,2011 [53]. For detailed mapping study see Appendix 1. Initial search by applying general string at selected databases for mapping study was:

Patient AND Data AND (Privacy OR Security)

This string retrieved 4,670 references. After screening, 58 studies were selected. There are many diversified terms used to address EHRs for patients in literature. It was a challenge to generate a valid string for targeting relevant studies. Therefore, we used the major terms of our selected primary studies from our mapping study to formalize a search string for our SLR. As a result, the following search string was produced:-

((Health Records" OR "Patient Information" OR "Medical Records" OR" Electronic Healthcare Records" OR "Electronic Health Records" OR" Electronic Medical Records" OR" Patient Data" OR" Medical Data" OR" Protected Health Information" OR "Personal Health Records") AND (Privacy OR Security))

3.1.1.3 Research Questions

As mentioned, the results of mapping study confirmed the feasibility of research questions phrased for SLR. The most important stage in SLR is to phrase relevant questions. In introduction already described importance of the research questions. Furthermore, to clarify reliability, mapped these questions against PICOC (Population, Intervention, Context, Outcome and Comparison) criteria [11].

RQ.1 Which solutions of patient data privacy and security have been proposed for software system development?

Population: Empirical Literature for solutions of patient data privacy and security.

Intervention: Framework, method, methodology, process, model etc.

Context: Healthcare sector.

Outcome: List of empirical solutions proposed for patient data privacy and security.

Comparison: No.

RQ.2*. Can we categorize these solutions against privacy and security principles proposed by Markel foundation [10]?*

Population: List of empirical solutions on patient data privacy and security for software development.

Intervention: Framework, method, methodology, process, model etc.

Context: Healthcare sector.

Outcome: Solutions for patient data privacy and security propose for software system development to answer privacy policy principles.

Comparison: Solutions to answer privacy and security principles – Stages of software system development on the basis principles proposed by Markel foundation [10].

The aforementioned research questions were intended to "assess the effect of software engineering technology" [11]. Whereas, 2nd research question will compare the population of intervention [11] to address specific issue of patient data privacy and security. Table 1 shows refinements of research questions.

Status	Date	Research	Issue
Initial Research questions	April 2010	***RQ1:*** *What are commonly existing techniques /methodologies/processes/models used to address issues related to patient privacy requirement engineering?* ***RQ2:*** *What are the gaps that have been reported in the proposed solutions identified in RQ1?*	➢ Is RQ1 able to retrieve solutions from literature? ➢ Which gaps reported in literature or have to perform analysis? ➢ Used common

		***RQ3**: Is there any framework that can be used for classifying reported gaps and the identified solutions in RQ2?*	word for solutions
Revision 1	June 2010	***RQ1:** What type of Patient data privacy solutions can be identified from the existing literature?* ***RQ2**: How these limitations or gaps of Solution from RQ1 can be classified to show current state of practice?*	➢ Literature reported large amount of solutions on patient data privacy.
Revision 2	Aug 2010	***RQ1:** What type of Patient data privacy and security solutions can be identified from the existing literature?* ***RQ2**: How these limitations or gaps of Solution from RQ1 can be classified to show current state of practice?*	➢ How reported solutions will be classified?
Revision 3	January 2011	***RQ.1** Which solutions of patient data privacy and security have been proposed for software system development?* ***RQ.2**. Can we categorize these solutions against privacy and security principles proposed by Markel foundation [10]?*	➢ Is it feasible to classify solutions against privacy and security principles of Markel foundation?
Final Research Questions.	April 2011	***RQ.1** Which solutions of patient data privacy and security have been proposed for software system development?* ***RQ.2**. Can we categorize these solutions against privacy and security principles proposed by Markel foundation [10]?*	------------------------

Table 1 Refinement of Research Questions

3.1.1.4 Reviewed Protocol

A detailed protocol have been developed [11]. The purpose of developing a protocol is to reduce bias and repetition of SLR procedure in auditable manner. The protocol is develop to perform review which, specifies the background for the review, research questions, search strategy, inclusion criteria, exclusion criteria, quality criteria, data extraction and methods to synthesized the extracted data. The protocol has been evaluated in two phases: (i) internal evaluation (author individually executed queries and extracted data by following the procedure of protocol); and (ii) external evaluation by the experts. (See detailed protocol in Appendix 1). For expert opinion, author requested three experts to review protocol. Table 2 shows valuable feedback from three experts. The names of those experts are:

- Barbara Kitchenham, Professor, Keele University UK.
- Dr. Mehmood Niazi, Lecturer, Keele University UK.
- Dr. Naveed Ikram, Associate Professor, Riphah International University Islamabad, Pakistan.

Parts of Protocol	Comments	Refined protocol
The Topic of Thesis	BK: Nil MN: Relevant NI: Perform mapping study	Perform mapping study
Search Strings	BK: Reviewing your search strings - I don't think "British Medical Society" is a good string - its a source not a keyword. I think you should include more relevant keywords from known set of papers. MN: It is not clear how the keywords and synonyms are identified.	Everyone
Search Sources	BK: You have a large number of sources and a large number of search strings. This can get complicated - even with Endnote it will take time to find duplicate papers because different sources have different standards for reporting authors names (some start with given names, some start with family names) and for referencing conference papers (sometimes the conference is used, sometimes the proceedings names are used MN: Nil	Everyone
Inclusion/Exclusion Criteria	BK: Be more specific about inclusion & exclusion criteria. Are you concentrating only on theoretical papers or are you including papers that report empirical trials of the proposed standards. Such	Everyone

	papers would need to be treated differently for example empirical papers might identify gaps in previous theoetical papers. What about papers that address privacy requirements for persons with data in multiple (perhaps heterogeneous) data sources in general without specifying the word "patient". Your inclusion/exclusion criteria (partyiculalry exclusion) are important for assessing the limitations of your results. MN: Not clear	
Data Extraction Form	BK: You need to consider how you are going to answer your questions - find some studies (with a quick search of one source) and check that you can extract information to address your questions - you need a data extraction form and trying out your data extraction process is a way to develop/prototype such a form. MN: Nil	Everyone
Research questions	BK: The origin of the "Gaps" information in the data extarction is not clear - do these arise from limitations reported in the paper or from your assessment of the paper? If you are reporting both sets of gaps you need to report each separately. addressing the question how does the data you collect answer the research questions? You also need to make clear the relationshp between research questions, inclusion/exclusion data extraction and data aggregation. The first two questions can be obtained from theoretical papers that discuss approaches. However, there may be overlapping suggestions - how do you intend to categories the different approaches to identify the set of unique approaches and within each approach categories the variety of different methods/techniques. The third question may be addressed by several different methods (as discussed above) - looking for empirical papers that critique certain approaches, identifying limitations reported by the authors of theoretical papers, critically assessing each paper yourself. You need to specify which approach you are using and make the information obtained from each source clear to the reader of the systematic review (this is required to address the auditability/traceability requirement of the systematic review methodology). MN: Nil	Everyone
Selection of primary studies	BK: Nil MN: Nil	Everyone
Motivation part	BK:Topic is very interesting and you have defended it well. MN: Good defense	------------------------- -

PICOC scheme	BK: Nil MN: Nil	-------------------------
Primary study selection process:	BK: Nil MN: Nil	-------------------------
Quality assessment	BK: Nil MN: Nil	-------------------------
Data extraction strategy and synthesis of the extracted data	BK: Nil MN: Nil	-------------------------
Synthesis of the extracted data	BK: Nil MN: Nil	-------------------------
Dissemination	BK: Nil MN: Nil	-------------------------

Table 2 Feedback from Experts

3.1.2 Conducting

This section describes detail procedure used for conducting SLR.

3.1.2.1 Piloting

Two separate search strings were run which gave lots of paper but time doesn't allow refining all of them only 10 are selected to validate the protocol the purpose of it not to show completeness but consistency in results and validate the protocol in auditable and repeatable manner.

Consistency in results

Papers retrieved from database: 2 different combination of research terms are used in IEEE explorer.

TITLE-ABSTR-KEY (("Health Records" OR" Patient Information" OR "Medical Records" OR" Electronic Healthcare Records" OR" Electronic Health Records" OR" Electronic Medical Records" OR" Protected Health Information" OR" Healthcare Records" OR "Health Record System")) and TITLE-ABSTR-KEY ((Privacy OR Security))

TITLE-ABSTR-KEY (("Healthcare Records" OR "Health Record System")) and TITLE-ABSTR-KEY ((Privacy OR Security))

Result: It is important that this category is the same for author and fellow researcher.

There is a close to 100% agreement on the papers downloaded from the database using the given search terms.

All references Library: All papers downloaded from database in this library, not pass full inclusion and exclusion criteria these all will be not included in final review.

Result: There is a difference in numbers of papers downloaded well all of these will be not included in final review. Differences pointed out in 'paper _accepted' library.

Work in Progress (WIP) library: This library is a temporary Store for papers that need more information (e.g. full papers) before a decision can be made.

Result: There are 2 papers initially stored in this library when discussion made both are deleted.

Papers that meet the inclusion criteria: Papers that meet inclusion criteria (answers a RQ, reliable source)

Result: There was only one paper difference.

3.1.2.2 Search Strategy

On the suggestion of our 1st expert (see acknowledgments), avoided manual search and reduced the number of databases. The objective was to simplify the search and reduce the time involved without compromising on the coverage of the results. Therefore, IEEE Xplorer, ACM Digital library, Science Direct and Springer Link were searched with the time starting from 2000 up to three quarters of 2011. Total 633 references were downloaded with abstracts and keywords in Endnote library [54] from 4 resources; *(f=367)* from IEEE Xplorer, *(f=72)* ACM Digital Library, *(f=115)* Science Direct and *(f=79)* Springer Link. The title, abstract and keywords were used as search identifiers. Author selected 123 studies from 633 records. However, 77 independent studies were included in the review. We also check reliability of search string. The string has given 100% coverage against already selected set of papers from mapping study. Table 3 shows lookup selection of primary studies from four databases.

Date	Search String TI=within title |AB=within Abstract| All= within ALL , Keywords| KW	Search Identifier	Comment		
15th Sept,2011	((Health Records" OR" Patient Information" OR "Medical Records" OR" Electronic Healthcare Records" OR" Electronic Health Records" OR" Electronic Medical Records" OR" Patient Data" OR" Medical Data" OR" Protected Health Information") AND (Privacy OR Security))	TI and AB	RQ1-search		
		1 separate searches (2000-2011)			
		Databases	**ALL**	**TI, AB, KW**	**Final**
		IEEE	367	66	46
		ACM Library	72	19	5
		Science Direct	115	31	21
		Springer Link	79	7	5
		Total	**633**	**123**	**77**

Table 3 Primary studies selection from Database

3.1.2.3 Inclusion and exclusion criteria

This review only included those empirical published studies, which have been peer-reviewed in journals and conferences published during the time span from 2000 up to three quarters of 2011. This inclusion criterion was based upon the evidence provided by mapping study [53]. The studies referring to any of the aforementioned privacy and security principles [10] for any phase of software system development were part of our review.

Those studies not explicitly providing solution for privacy and security of EHRs or supporting any other area of software engineering rather than software system development and not published in English language were excluded. We also excluded those studies that proposed solutions for privacy and security of patient images, wireless communication, sensor networks and in home monitoring systems. We also excluded books, technical reports, project thesis studies based on philosophical research and expert opinions. Figure 3 shows total number of studies selection at different stages.

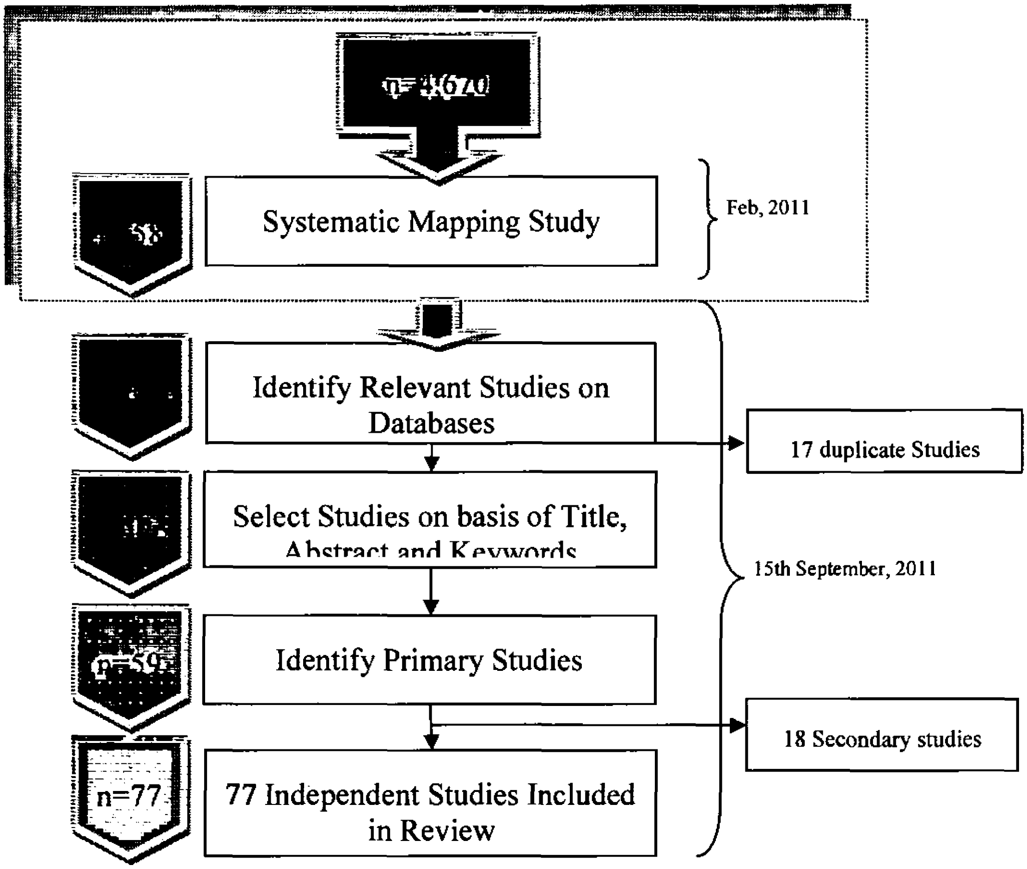
n=3670
Systematic Mapping Study
Feb, 2011
Identify Relevant Studies on Databases
17 duplicate Studies
Select Studies on basis of Title, Abstract and Keywords
15th September, 2011
n=59
Identify Primary Studies
18 Secondary studies
n=77
77 Independent Studies Included in Review

Figure 3 Studies selection at different stages

3.1.2.4 Data extraction

At this stage of conducting phase data of selected primary studies from previous phase extracted. To carry out data extraction more efficiently forms were designed in MS word. These forms also help in consistency of data extraction. In piloting of protocol forms were already evaluated. It is difficult to set values of all properties prior to data extraction. These properties are totally dependent to papers and their contents. However, the extracted properties with relevant questions are mentioned in Table 4.

ID	Property	Research questions
P1	Solution type	RQ1
P2	Software development lifecycle	RQ1
P3	Privacy and security principles	RQ2
P4	Research Type	Overview of the studies
P5	Context	Overview of the studies

Table 4 Extracted Properties

P1 was extracted on the basis of study's authors' terms like framework, method, methodology, model etc. P2 was assigned on predefined set of values based upon waterfall model for software development lifecycle like requirement engineering, software architecture and design and implementation, etc. As mentioned before, 9 privacy and security principles of Markel foundation [10] were used for categorization of proposed solutions, P3 was assigned to these values. Forth property (research type) each study classified according to the research type categorization proposed by Wieringa R [55]. This categorization clarifies the research facet with empirical evidence in the concerned area. Table 5 describes each research type. P5 was used to assigned values based upon context/environment of the proposed solution.

Category	Description
Validation Research	"Techniques investigated are novel and have not yet been implemented in practice. Techniques used are for example experiments, i.e., work done in the lab".
Evaluation Research	"Techniques are implemented in practice and an evaluation of the technique is conducted. That means, it is shown how the technique is implemented in practice

	(solution implementation) and what are the consequences of the implementation in terms of benefits and drawbacks (implementation evaluation). This also includes identifying problems in industry."
Solution Proposal	"A solution for a problem is proposed, the solution can be either novel or a significant extension of an existing technique. The potential benefits and the applicability of the solution is shown by a small example or a good line of argumentation".
Philosophical Papers	"These papers sketch a new way of looking at existing things by structuring the field in form of a taxonomy or conceptual framework".
Opinion Papers	"These papers express the personal opinion of somebody whether a certain technique is good or bad, or how things should been done. They do not rely on related work and research methodologies".
Experience Papers	"Experience papers explain on what and how something has been done in practice. It has to be the personal experience of the author".

Table 5 Research Type Categorization

3.1.2.5 Study Quality Assessment

All independent 77 studies were selected at stage 4 and 5 as shown in Figure 6. These studies were critically assessed by quality guidelines which we adopted from [56]. We have adopted very structural approach to validate the records of data extraction and quality assessment of the studies. Along with author, 2 more experts validated quality assessment of the studies. Selected studies were divided in two categorizes qualitative and quantitative. In quantitative studies, proposed solutions were based upon rigorous and credible results of experimentation expressed in frequencies. While, qualitative studies results were usually sensitive, detailed and contextual. However, 84% of our studies were qualitative and 16% were quantitative. Each question appeared, in our quality assessment checklist as shown in Table 6 was answered as (Yes=1, No=0, partially=0.5). Quality score for quantitative study lie within a spectrum of 0 to 19, and for a qualitative study lie within a spectrum of 0 to 8. No single study got 100% score. Table 12 shows or quality assessment checklist. The results of quality assessment will be given in subsequent sections.

Qualitative Guidelines	
Sr no.	Questions
1	Is the research design suitable for carrying out the study?
2	Does the study build upon existing body of knowledge, i.e., does it explicitly discuss its contribution in the light of previous work?
3	Does the study report clear, unambiguous findings based on evidence and argument?
4	Are the findings credible?
5	Is the research process described thoroughly? Are the roadblocks, false steps described in a helpful way?
6	Are the links between the data, interpretation, and conclusions clear?
7	Is the reporting clear and coherent?
8	Are the assumptions/theoretical perspectives/values that have shaped the form and output of the evaluation clear?
Quantitative Guidelines	
1	Are the research question(s) clearly stated for the studies?
2	Does the study build upon existing body of knowledge, i.e., does it explicitly discuss its contribution in the light of previous work?
3	Are the variables/metrics used in the study adequately measured and validated?
4	Are the metrics used in the study clearly defined?
5	Are all model construction methods/metric(s) derivation methods fully defined (tools and methods used)?
6	Are the metrics used in the study the most relevant ones for answering the research questions?
7	Are the data collection methods adequately described?
8	Are the statistical methods justified?
9	Is the purpose of the data analysis clear?
10	Are potential confounders adequately controlled in the analysis?
11	Are the negative findings presented?
12	Do the researchers discuss any problems with the validity/reliability of their results?

13	Is the study replicable?
14	Is the research design clearly presented?
15	Is the research design suitable for carrying out the study?
16	Are the findings credible?
17	Is the research process described thoroughly? Are the roadblocks, false steps described in a helpful way?
18	Are the links between the data, interpretation, and conclusions clear?
19	Is the reporting clear and coherent?

Table 6 Quality Assessment Criteria

3.1.2.6 Data Synthesis

"Data synthesis involves collecting and summarizing the results of the included primary studies. Synthesis can be descriptive (non-quantitative). However, it is sometimes possible to complement a descriptive synthesis with a quantitative summary." [11]. The extracted data from data extraction forms were recorded on excel sheets. This really helped us to find trends, consistency and relevant similarities for analysis of data. We describe analysis of the results in subsequent chapter.

3.1.3 Documenting

This Systematic literature review reported as a final thesis of MS(SE) degree. The results of our mapping studies already published in 2011 [53] and short version of this thesis also submitted for journal publication.

3.1.3.1 Validity Threats

As suggested by Kitchenham [11] validity threats like from Investigator Bias, Publication Bias, and Data Extraction Consistency threats also considered during SLR. The detail description of this phase will be explained in Chapter 4.

4 Results and Analysis

This chapter describes the result gained by data extraction, Step 8 of SLR. Meanwhile, these results have been analyzed and discussed. First a brief overview of the selected primary studies and their general characteristics are discussed. This is followed by the results and analysis associated to research questions.

4.1 Overview of studies

From 123 studies a total of 46 studies did not meet the minimum selection criteria. Out of these, 9 studies were proposed privacy and security for biomedical images, 6 studies were extension of already included studies of the same authors, 7 were reporting privacy and security for sensors network while, 8 focused on in-home monitoring patients. In the first cycle, data of 59 independent studies were extracted in electronic forms. In this cycle 31 new references were recorded. In the second cycle, 18 new independent studies were included in the review, while 7 were duplication and 6 were out of the scope (see Figure 3).

4.1.1 Publication Channels

The selected 77 independent studies were published in 51 different channels, as shown in . Most of the studies were published in conferences (46 out of 77, 59%), whereas others (33 out of 77, 42%) appeared in journals. The top publication channels were two journals; International Journal of Medical Informatics (21.5%) and Information Technology in Biomedicine (5%). The complete distribution of studies against each publication channel is shown in Table 7.

Publication Channels	Type	Number	Percent
International Journal of Medical Informatics	Journal	16	20.7
Information Technology in Biomedicine	Journal	4	5
CBMS	Conference	3	3.8
Information Technology Applications in Biomedicine	Conference	3	3.8
D2H2 2006	Conference	2	2.5
ARES 08	Conference	2	2.5
ICIT 2006	Conference	2	2.5

Pervasive Computing Technologies for Healthcare 2008	Conference	2	2.5
POLICY	Conference	1	2.5
ACM SIGKDD	Conference	1	1.2
BioCAS 2006	Conference	1	1.2
ICACT 2008	Conference	1	1.2
PRDC 2007	Conference	1	1.2
ACM SE '10	Conference	1	1.2
ACT '09	Conference	1	1.2
AMS 2010	Conference	1	1.2
Artificial Intelligence and Law	Journal	1	1.2
Artificial Intelligence in Medicine	Journal	1	1.2
ASIACCS '11	Conference	1	1.2
Communications and Multimedia Security	Journal	1	1.2
computer methods and programs in biomedicine	Journal	1	1.2
Data & Knowledge Engineering	Journal	1	1.2
Digital Information Management 2006	Conference	1	1.2
EDOC '08	Conference	1	1.2
E-health Networking, Applications and Services 2008.	Conference	1	1.2
HICSS '06	Conference	1	1.2
ICBECS 2010	Conference	1	1.2
ICDE 2005	Conference	1	1.2
ICITIS 2010	Conference	1	1.2
IMIS 2011	Conference	1	1.2
INC 2010	Conference	1	1.2

INDIN 2003	Conference	1	1.2
INFOS 2010	Conference	1	1.2
IPC 2007	Conference	1	1.2
ISCE 2011	Conference	1	1.2
ITAB 2010	Conference	1	1.2
ITIME '09	Conference	1	1.2
Journal of Biomedical Informatics	Journal	1	1.2
Journal of Research and Practice in Information Technology	Journal	1	1.2
Knowledge and Information Systems	Journal	1	1.2
NSS '09	Conference	1	1.2
Parallel and Distributed Systems	Journal	1	1.2
PRIVACY, SECURITY AND TRUST 2005	Conference	1	1.2
Requirements Engineering	Journal	1	1.2
Security & Privacy, IEEE	Journal	1	1.2
SIGMOD 2000	Conference	1	1.2
Software Engineering	Journal	1	1.2
System Sciences	Conference	1	1.2
VLDB '02	Conference	1	1.2
WCMeB 2007	Conference	1	1.2
WD 2008	Conference	1	1.2
Total	**51**	**77**	**100**

Table 7 List of Publication Channels

We also examined the citation status of these studies from Google Scholar [57]. Here, we found that 2 studies [20, 58] were cited more than 500 times, 4 studies [31, 59-61] were cited less than

100 times, while sixty studies were cited less than 50 times e.g. [27, 29, 30, 32, 36, 37, 61-85]. This gives an indication to the quality of our primary studies.

4.1.2 Publication Years and Research Community

Figure 4 shows the trend of publications, year wise. Here, we observed that quite low number of studies was published in the years before 2005, but the publication rate increased after the year 2005. We noticed more publications in 2008 *(f=13)*, 2010 *(f=12)*, up to three quarters of 2011 *(f=10)* and 2009 *(f=8)*. However, 2000 *(f=1)* and 2001 *(f=0)* were found with least publication rate. The countries which are most active in this area of research are USA, Australia, UK and Canada. Figure 5 gives an overview of the countries along with its frequencies of publications.

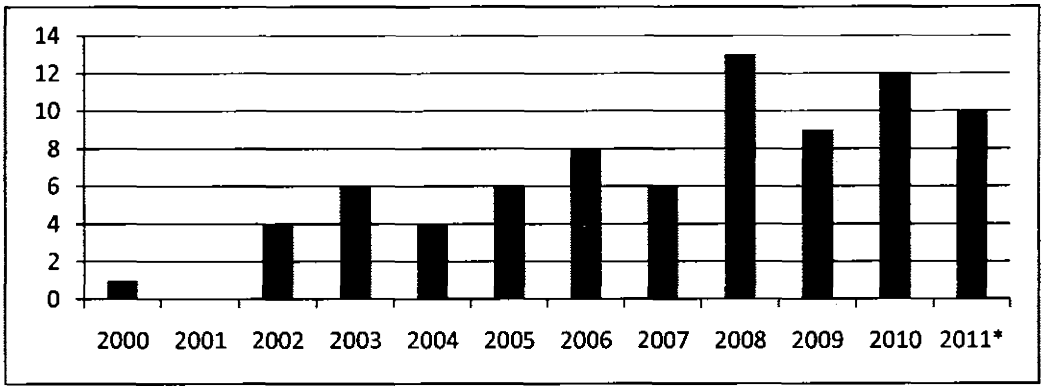

Figure 4 Distribution of Papers According to Years

* Three quarters of 2011.

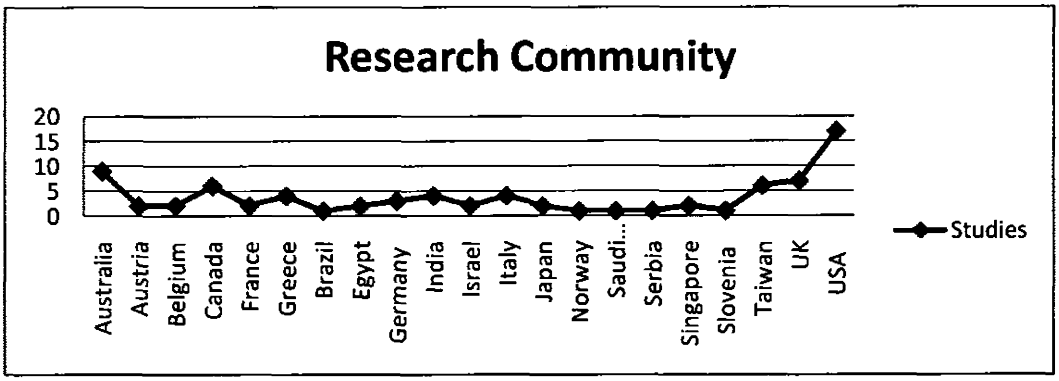

Figure 5 Overview of Research Community

4.1.3 Studies Quality Assessment

The quality of 77 independent studies were assessed on the basis of credibility and relevance against predefined quality guidelines [56]. Author independently assessed the quality of each study after reading whole paper. Among the selected studies, 84% studies were qualitative and 16% were quantitative. We used a quality assessment checklist [56] to quantify the quality of study. Each of the question on the checklist was answered as (Yes=1, No=0, partially=0.5). Table 8 shows the quality assessment checklist. In the qualitative studies, a qualifying score was between (7.5 - 6 out of 8). The studies falling in this group did not define their contribution in existing body of knowledge or have weak evaluation of output. The studies scoring (5.5 - 4 out of 8) did not had a relationship between data interpretation and its conclusion. The qualifying score for quantitative studies was between (16 - 9.5 out of 18). The qualifying score of quantitative studies lacks, unable to justify statistical method and didn't provided negative findings. Table 14 shows number of studies against each quality assessment guideline. Similarly, Figures 7 and 8 illustrate frequency of assessment results.

Qualitative Guidelines						
Questions	**Yes**	**Number**	**Partially**	**Number**	**No**	**Number**
1. Is the research design suitable for carrying out the study?	[86, 87][20, 22-24, 27, 30, 33-38, 58-62, 64-68, 70, 72, 74, 76, 78-85, 88-105]	34	[32, 69, 71, 73, 75, 77, 106-110]	12	[29]	1
2. Does the study build upon existing body of knowledge, i.e., does it explicitly discuss its contribution in the light of previous work?	[20, 23, 27, 30, 33, 34, 36-38, 58, 59, 61, 64, 66, 67, 69-73, 76, 79, 80, 83, 84, 86, 88-94, 96-98, 100, 102, 103, 105, 106]	40	[22, 24, 29, 32, 35, 60, 62, 65, 68, 74, 75, 77, 78, 81, 82, 85, 87, 95, 99, 101, 104, 107-110]	25		1

3. Does the study report clear, unambiguous findings based on evidence and argument?	[20, 22-24, 27, 32-38, 58, 60-62, 64-66, 68, 69, 72, 74, 76, 78-80, 82, 83, 85-96, 98-106, 108-110]	44	[30, 59, 67, 70, 71, 73, 77, 84, 97, 107]	10	[29, 75, 81]	3
4. Are the findings credible?	[22-24, 27, 29, 30, 32-34, 36-38, 60-62, 64-66, 69, 71-78, 80-100, 102-110]	48	[20, 35, 58, 67, 68, 70, 101]	7	[59]	1
5. Is the research process described thoroughly? Are the roadblocks, false steps described in a helpful way?	[33, 36, 60, 67, 77, 82, 83, 88, 98, 102, 103]	11	[20, 22-24, 27, 30, 32, 34, 35, 37, 38, 58, 59, 61, 62, 64-66, 68-70, 72-74, 76, 78-81, 84-87, 89-94, 96-101, 104, 105, 109]	49	[29, 71, 75, 95, 106-108, 110]	8
6. Are the links between the data, interpretation, and conclusions clear?	[20, 22-24, 27, 30, 32-35, 37, 38, 58-62, 64, 65, 67-70, 72, 74, 76, 77, 80-87, 89-96, 98-101, 103-109]	55	[36, 66, 71, 73, 78, 79, 88, 102, 110]	9	[29, 97]	2
7. Is the reporting clear and coherent?	[20, 22, 23, 27, 30, 32, 34-38, 58, 60-62, 64-70, 72-74, 76-90, 92, 94, 95, 98-100, 103-109]	53	[24, 33, 59, 71, 91, 93, 97, 101, 102, 110]	10	[29, 96]	2
8. Are the assumptions/theoretical perspectives/val	[20, 22-24, 27, 32-34, 36-38, 58, 60, 62, 65,	41	[29, 30, 35, 59, 61, 64, 66, 68, 69, 71, 77, 78,	23	[106]	1

ues that have shaped the form and output of the evaluation clear?	67, 70, 72-74, 76, 79, 81, 82, 84, 85, 87, 89-97, 99, 101, 102, 109, 110]		80, 83, 86, 88, 98, 100, 103-105, 107, 108]			
Total Frequency	**326**		**145**		**19**	
Quantitative Guidelines						
1. Are the research question(s) clearly stated for the studies?	[26, 31, 63, 111-115]	8	[25]	1		
2. Does the study build upon existing body of knowledge, i.e., does it explicitly discuss its contribution in the light of previous work?	[26, 31, 63, 111, 112, 114]	6	[25, 113, 115]	3		
3. Are the variables/metrics used in the study adequately measured and validated?	[26, 63, 111, 112, 114, 115]	6	[31, 113]	2	[25]	1
4. Are the metrics used in the study clearly defined?	[26, 63, 111, 112, 114]	5	[31, 113, 115]	3	[25]	1
5. Are all model construction methods/metric(s) derivation methods fully defined (tools and methods used)?	[63, 111]	2	[25, 114, 115]	2	[26, 112, 113]	3
6. Are the metrics used in the study the most relevant ones for answering the	[63, 111, 112, 114, 115]	4	[31, 113]	2	[25, 26]	2

research questions?						
7. Are the data collection methods adequately described?	[31, 63, 111-115]	7	[26]	1	[25]	1
8. Are the statistical methods justified?	[26, 31, 63, 111]	4	[113, 115]	2	[25, 63, 112, 114]	4
9. Is the purpose of the data analysis clear?	[26, 63, 111-115]	6	[25, 31]	2		
10. Are potential confounders adequately controlled in the analysis?	[26, 31, 63, 114]	4	[111-113]	3	[25, 115]	2
11. Are the negative findings presented?	[63, 115]	2	[31, 113, 114]	3	[25, 26, 111, 112]	4
12. Do the researchers discuss any problems with the validity/reliability of their results?	[26, 31, 63, 115]	4	[113]	1	[25, 111, 112, 114]	4
13. Is the study replicable?	[25, 26, 111, 113, 114]	5	[31, 63, 112]	3		
14. Is the research design clearly presented?	[63, 111, 112, 114]	4	[25, 113]	2	[26, 31, 115]	3
15. Is the research design suitable for carrying out the study?	[31, 63, 112, 113, 115]	5	[25, 111, 114]	3	[26]	1
16. Are the findings credible?	[25, 26, 31, 63, 111, 115]	6	[113, 114]	2		
17. Is the research process described	[31, 63, 113-115]	5	[26, 111, 112]	2	[25]	1

thoroughly? Are the roadblocks, false steps described in a helpful way?						
18. Are the links between the data, interpretation, and conclusions clear?	[26, 31, 63, 111-115]	8			[25]	1
19. Is the reporting clear and coherent?	[26, 31, 63, 111-115]	8	[25]	1		
Total Frequency	**98**		**44**		**30**	

Table 8 Quality Assessment Guidelines and Results

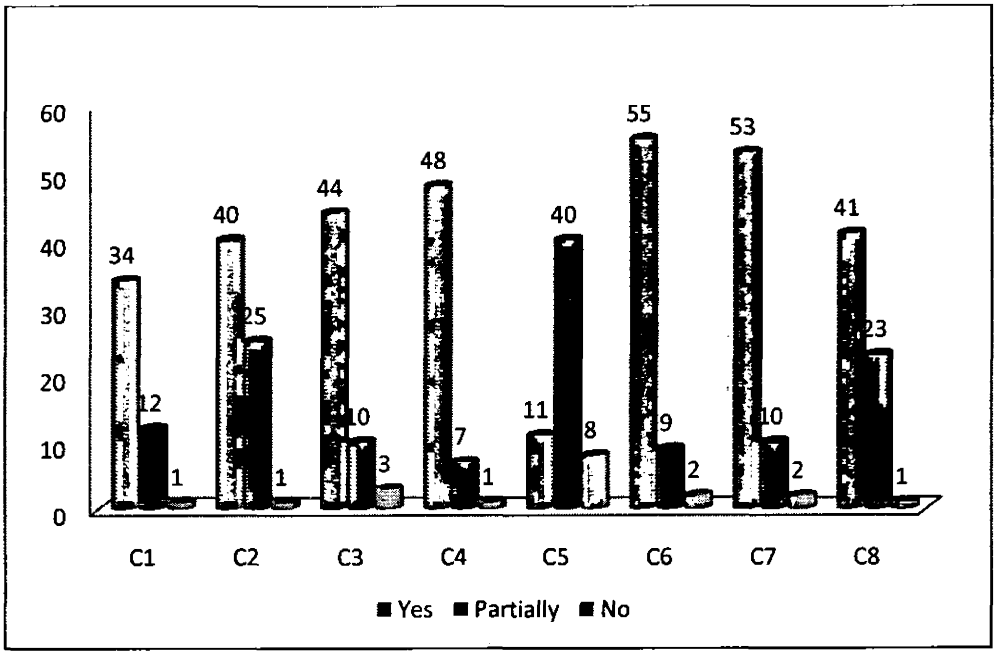

Figure 6 Qualitative Assessment Results

C1, C2, C3.....,C8 Qualitative guidelines

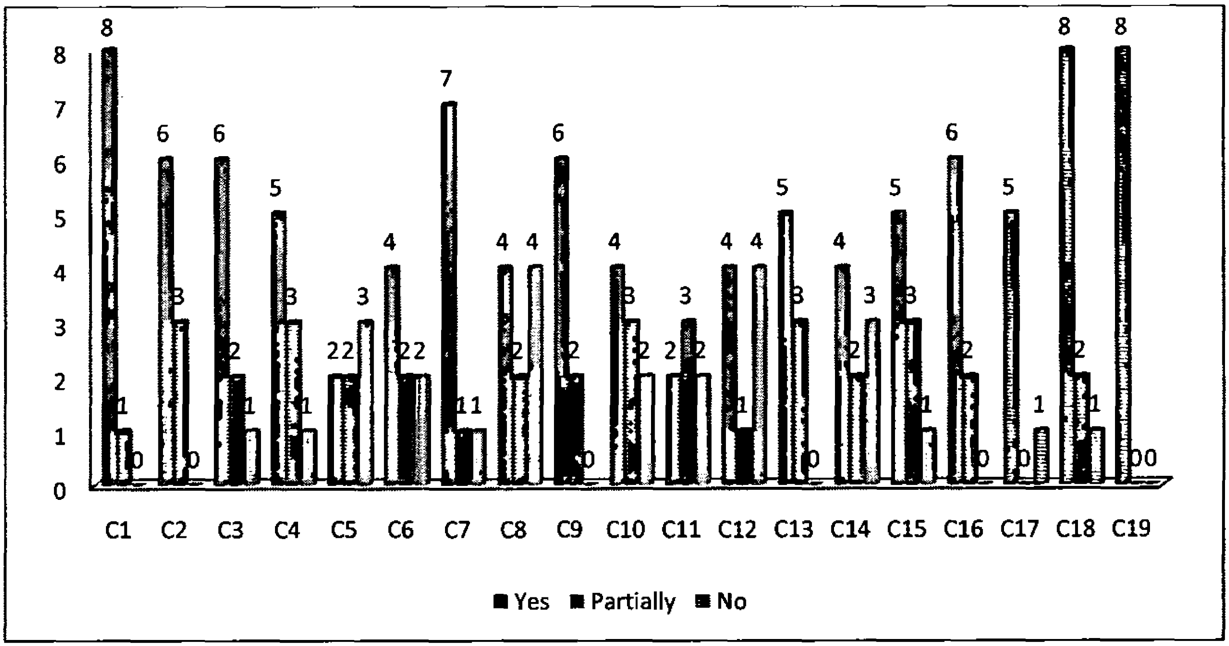

Figure 7 Quantitative Assessment Results

C1, C2, C3.....,C19 Quantitative guidelines

4.2 Solution Type (P1) RQ1

This property is extracted to show the evidence and types of solution proposed in literature to answer patient data privacy and security. Solution type names are based upon the study's author's terms. Studies offer some type of solution like method, framework, guideline or model that can be used by other researchers in this area. Among 77 papers, the most common solution type is approach (16 papers, 20.7%) followed by framework (14 papers, 18%) and least common solution types are process and guidelines (1 paper, 1.2%). Table 9 shows complete distribution of studies against each solution type with their frequencies.

Type of Solution	Papers	Frequency
Approach	[20, 21, 59, 64, 69, 70, 78, 82, 89, 96, 98, 100, 103, 106, 113, 114]	16
Framework	[22, 26, 31, 66, 68, 71, 76, 80, 84, 94, 102, 107, 110, 112]	14

Guidelines	[38]	1
Method	[21, 23, 61, 63, 72, 90, 93]	7
Methodology	[30, 60, 65, 88, 101]	5
Model	[25, 27, 29, 67, 73, 77, 92, 116]	8
Process	[99]	1
Scheme	[37, 38, 61, 74, 86, 90, 99, 108]	8
Technique	[32, 75, 81, 85, 91]	5
Not mentioned	[24, 33, 35, 58, 62, 75, 83, 84, 97, 105, 111, 117]	12

Table 9 Types of Proposed Solutions

4.3 Software Development Lifecycle (P2) RQ1

As our first research question of this thesis is to retrieve proposed solutions of patient data privacy and security for software system development. 77 independent studies are categorized under SDLC phases (Requirement Engineering, Software Architecture and Design, Implementation etc). This categorization clearly shows which SDLC phases are most and least research in this area and their effects on different software development process phases. The most investigated SDLC phase in this area is software architecture and design with 64%, followed by implementation 27%. However, literature do not report any study for software testing and maintenance phase. Figure 9 shows complete percentages of SDLC phases. Table 10 illustrates complete distribution of 77 studies according to SDLC phases.

SDLC Phases	Studies	Frequency
Requirement Engineering	[27, 30, 60, 68, 81, 87, 101]	7
Software Architecture and Design	[20-24, 26, 29, 32, 35, 37, 59, 62, 64, 65, 69, 71, 72, 74-80, 83, 84, 86, 88, 89, 93-100, 102, 105-110, 112, 113, 116]	47
Implementation	[25, 31, 33, 34, 36, 38, 58, 61, 63, 66, 67, 70, 73, 82, 85, 90-92, 103, 104, 111, 114, 115]	23

Table 10 Distribution of Studies According to SDLC Phases

These SDLC phases are further categorized according to privacy and security principles [10] in next section.

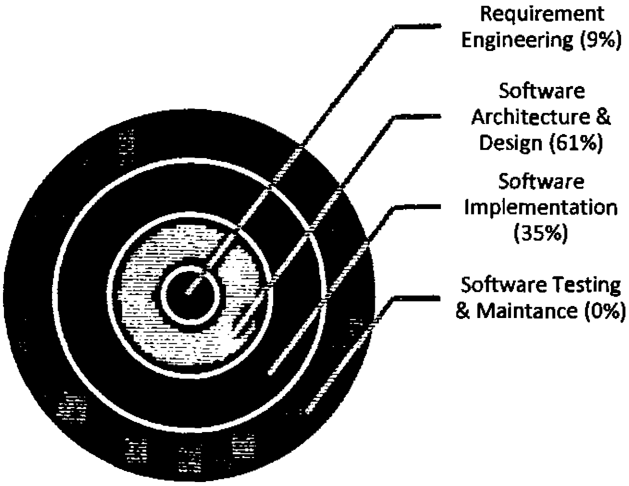

Figure 8 Percentage of SDLC phases

4.4 Privacy and Security Principles (P3) RQ2

Our second research question is categorization of patient data privacy and security solutions proposed for software system development against privacy and security principles of Markel foundation [10]. All nine principles [10] are covered in one or another spectrum. We analyze and discuss each proposed solution according to privacy and security principles. These nine principles are further categorized according to SDLC phases and retrieved solutions from review.

Table 11 shows distribution of 77 studies against each solution with frequencies. Table 12, 13 and 14 show details of solutions with their research types.

Privacy and Security Principles	Proposed Solutions	RE	F	SA & D	F	IMP	F	Total Frequency
Security Safeguard and Control	Authentication	[30]	1					1
	Access Control	[30, 60, 68, 81, 101]	5					5
	Secure Data Transmission	[81, 87]	2					2
	Smart Card and Biometrics			[29, 65, 71, 78]	4			4
	Encryption			[25, 32, 91, 106]	4	[36, 108]	2	6
	Pseudomization			[20, 21, 77, 88, 89, 93]	5	[38]	1	6
	Establish Trust			[23, 116]	2			2
	Ontology Based Technology			[31]	1			1
	Patient defined Access Condition			[62]	1			1
	k-anonymization			[95, 96, 112]	3			3
	Web-services			[64, 97, 107]	3			3
	Behavior based Access Control			[34]	1			1
	Role based Access Control			[59, 72, 74, 76, 80, 100, 102, 104, 105]	9			9
	Situation based Access control			[83]	1			1

	Authentication Certificate			[35, 98]	2			2
	Firewall			[86, 110]	2			2
	Emergency Access Token			[79]	1			1
	Cryptography			[69, 94, 99, 109]	4			4
	Anonymization					[70, 90, 114]	3	3
	Exponential and Laplace Noise					[63]	1	1
	Task Independent Technique					[85]	1	1
	Quantified Risk-adaptive Approach					[115]	1	1
	Context Specific Scheme					[37]	1	1
	Key Management Scheme					[33]	1	1
	Decision Learning Tree					[58]	1	1
	Secure Dedup Algorithm					[111]	1	1
	Bipolar Multiple Number Base Technique					[61]	1	1
	Data and Data Structure Modification					[66]	1	1
Accountability and Oversight	Oversight	[68, 74]	2					2
	Traceability	[81]	1	[95]	1			1

	Auditing			[20, 37, 59, 68, 74, 77, 84, 100]	8			8
Purpose Specification and Minimization	Minimum Information collection	[30]	1					1
	Compliance with Law	[87]	1					1
	Patient Defined Purpose			[20]	1			1
	Purpose based Collection					[82]	1	1
Individual Participation and control	Data ownership	[68, 81]	2					2
	Patients defined access control rules			[20, 62, 72, 77]	4			4
	Pseudonymizatlon			[21, 88]	2			2
	ChiperMe Technology			[32]	1			1
	I-Keys Hardware			[29]	1			1
	Cryptographic Credentials			[69]	1			1
	Digital Rights Management			[35]	1			1
	Anonymization					[70]	1	1
	Encryption					[36, 108]	2	2
	Bipolar Multiple Number Base Technique					[61]	1	1
	Access Management			[110]	1			1
Data Integrity and Control	Secure Transmission	[81]	1					1

	Data Quality	[101]	1					1
	Accurate EHRs			[20]	1			1
	Signatures Scheme			[74, 91, 104]	3			3
	Asymmetric Fingerprinting Scheme			[77]	1			1
	Cryptographic Key Management Scheme					[36]	1	1
Remedies	Code of Ethics	[81, 101]	2					2
	Copyrights					[31]	1	1
Use Limitation	Secure Transmission	[81]	1					1
	Data Quality	[101]	1					1
	Signature Schemes			[74, 91, 104]	1			1
	Accurate EHRs			[20]	1			1
	Asymmetric Fingerprinting Scheme			[77]	1			1

	Cryptography Key Management Scheme					[36]	1	1
Collection Limitation	Consent and Disclosure Limitation			[20]	1	[27]	1	2
Openness and Transparency	Public Key and Attributes Certificate			[106]	1			1
	Access Management			[110]	1			1
	Patient Understanding					[36]	1	1

Table 11 Distribution of Studies against Privacy Principles

Requirement Engineering RE

Software Architecture and Design SA & D

Implementation IMP

4.4.1 Security Safeguards and control

This is the most researched security principal in our review. 76 out of 77 studies reported this important principle. Under this principle, 64.4% studies were related to Software Architecture and Design, 28% were related to Implementation and only 6.5% were related to Requirements Engineering.

4.4.1.1 Requirement Engineering

4.4.1.1.1 Access Control

Risk assessment methodology of Gritzalis et al [30] also proposed access control on the basis of Role Based Access Control Model (RBAC). Another methodology of Compagna et al [68] proposed for development of security and privacy patterns to address the issue of security and privacy requirements as per defined legislation. The data requester has to provide an authorization certificate to obtain the required EHR. The proposed methodology is evaluated by using stable semantic model. Similarly, methodology of Massey et al [101] proposed to evaluate security and privacy requirements against the set of laws to check the system's compliance. His methodology suggested iTrust software requirements which said that e-sessions must get expired after a pre-defined period and illegal access to systems for three times must disable the unauthorized user for 15 minutes. The methodology is proposed to comply software requirements with security policy. There is another methodology [60] proposed to give statement level coverage from legal requirements. This methodology deduces six types of access controls on the basis of defined rules. The results of this methodology proposed by applying coverage at HIPPA privacy law. Furthermore, technique of Jenson et al [81] to map legal requirements for sensitive health information also provided access control by checking authorization level.

4.4.1.1.2 Authentication

Gritzalis S et al [30] proposed risk assessment methodology for privacy of EHRs. This methodology is empirically set in real scenario and proposed technical guidelines for patient privacy and security. Methodology propose strong authentication by PKI-enables smart cards. Methodology also highlights authorization risks due to varied environments.

4.4.1.1.3 Secure Data Transmission

Two studies [81, 87] also proposed that EHRs also ensure transmission security on communication channel.

4.4.1.2 Software Architecture and Design

4.4.1.2.1 Role Base Access Control

An architecture for Italian rehabilitation centre has been presented [105]. To design this architecture users of the system are divided into workgroups according to organizational structure. Access control is given according to roles of these workgroup. The proposed architecture is design to manage increasing number of healthcare actors, objects and related access issues under temporal conditions. Similarly, Han S et al [76] introduces new framework for authorization and authentication for healthcare sector. This architecture proposed the authorization then authentication followed by role privileges. This will help to control the access with different roles. The proposed architecture helps to mange e-health systems in secure and flexible way. Furthermore, security of HIS a Client-Server architecture is proposed by Lien C-C et al [100]. This architecture specifies the security requirements through roles. Architectural components like *AccessControl* and *AuditTrial* are responsible to control accesses. The architecture proposed that Client-Server approach is helpful to integrate security requirements in HIS. Without making compromises on any privacy rule a security *framework* [102] for communication and data sharing of EHRs components inside the organization has been proposed. To ensure security management in large organizational system, he used Role-Based Access Control (RBAC) security model. This helped him in assigning the proper roles to users and managed the legal authorization in the system. The proposed framework allows communication and sharing of EHRs by preserving privacy. For de-centralized EMRs, an approach has been proposed by Blobel [59] which separated structural roles using organizational entity-to-entity relationship to assign specific roles and duties to each role. This separation of roles helps in management of authorization. The proposed approach provides access control for EHRs in large organizations. A combination of Grounded Theory (Gt) with mixed methods can be used to involve health professionals in design and enhancement of access control of de-centralized EHRs [72]. He suggested that *Access Control Roles (ACRs)* must be thoroughly analyzed before they are defined. He also suggested that *only* temporary access control should be given to users in emergency cases. Gritzalis D et al. in 2004 [74] proposed authorization and authentication architecture based scheme for web based distributed systems. According to this scheme SSL/TLS are used for exchange of encrypted data between EHR servers. whereas, Role based access control(RBAC) model is followed for access control with single sign-on authentication and smart authorization services. The proposed architecture supports authorization and authentication in multiple policies domains. Another RBAC based framework proposed by Hung et al [80]. Authors extends existing framework to tackle privacy needs of e-Healthcare Services. To control user access on EHRs, RBAC is represented as a set of conditions of use declaration in context of users, roles, organizations, operations and datasets. The request received at access control entertained on basis of purpose, recipient, obligation and retention. There is another secure EHR system proposed [104] for patient data sharing in cross domain by preserving privacy. In this system fine grained access control over Public Key encrypted (PEKS) EHRs with proper assignment of roles.

4.4.1.2.2 Pseudomization

Patients self-determination model of Hass et al [77] entertains every access request by the permission of policy service. This model utilizes digital watermarking service (WMS) with pseudonymity service (PS) to make EHRs secure. The proposed model claims control access to EHRs by establishing patient profile. Neubauer T et al [88] also suggested a PIPE methodology to protect EHRs from unauthorized access under patient control. This methodology decoupled patient's identity with EHRs. There is an encrypted link between identification and health pseudonyms. Through Pseudonymization service, generalized EHRs are realized. This novel methodology allows secondary use of EHRs by preserving privacy. Similarly, strawman Design [20] proposed for Hippocrates databases which aim to provide proper authorization of EMRs. To guard the data from sneaking, this approach encrypted some data items. In another study, Quantin et al [93] proposed MRSE as an alternate to centralized HIS, which is based on pseudonymisation of patient identity. In this system the patient privacy is protected by pseudonymous code (based on patient's identity) with encrypted communications. The proposed system is implemented and proved as a better solution then centralized EHRs systems. To secure primary and secondary use of EHRs Riedl et al [89] proposed new PIPE architecture which guarantees privacy by access control in centralized environment. Instead of using traditional encryption the security model of this architecture used *pseudonymization* threshold scheme. The solution proposed guarantee patient privacy and security through pseudonymization with fallback mechanism. There is another architecture [21] for linking EHRs in a way that gives patients control over what information exposed about them . Indirect pseudonym identifiers are used for this purpose. A case study is used to show how architecture satisfies data accuracy needs and privacy requirements.

4.4.1.2.3 Smart Cards and Biometrics

A model [29] has been validated in Australian regional setting which provided authentication and authorization by utilizing I-Keys hardware. This model used smart card technology as secure method to access central EHRs. To validate model a field trial is made concerning 20 patients with their 6 General practitioners. Which results, mutually dependent I-Keys and smart tokens allow patients to award current access withdraw and transfer EHRs to authorized users. Similarly, Dwivedi A et al [71] proposed multilayer HIS framework based on combinations of

PKI (Public Key Infrastructure), Smartcard and Biometrics Technologies. These technologies are implemented in three security layers for access control against each stakeholder. The proposed solution highlights attention towards missing security technology in healthcare sector. Another secure approach presented by Hembroff C G et al [78] called SHIELD. This approach integrated in architecture to achieve security with scalability and interpretability. The approach utilizes smart card technology and biometrics to permit exchange of EHRs. This approach is implemented and working successfully running in fourteen hospitals. Furthermore, a methodology has been proposed [65] who used smart card technology to reach the required security level. This methodology assigned different access mode to each user according to his rights. They prevented the unauthorized usage through their memory protection mechanism. Further, to secure sensitive data, an encryption algorithm was chosen here. A methodology utilizes logical and physical databases design to support required level of security. There is another methodology using smart card technology [65] divided the entities in logical fragments (horizontally and vertically) in order to access need. There is a tag for each fragment with its cardinality. These fragments with need are allocated to the smart card. When this need match with information system record EHR is release for use. In this way access control on EHRs is mange by need.

4.4.1.2.4 Data Encryption

Hensen I [108] proposed ChiperMe architecture for secure storage of EHRs. This architecture neglects the concept of administrations and allows patients to control their own records. EHRs are selected and stored in ChiperMe server in encrypted form. In response of digitally signed request by access card ChiperMe server deliver encrypted EHR. The terminal decrypts the EHR by card for viewing and processing. Whereas, majority of the access tools and technologies are internet based. Furthermore, ChiperMe stores each EHR in separate locked boxes throughout storage and transportation. This approach proposed that ChiperMe application can be used to preserve privacy in medicine by empowering patients. A secure web-based confidential communication for heterogeneous EHRs is proposed by Choe [61]. In this method, a Client Application (CA) caters only those users who have a digital certificate. Then, CA sends request to Central-Access Control (CAC). To securely exchange this data between CA and CAC, XML encryption has been used. In 2007 Sucurovic S [106] proposed architecture for MEDIS (Medical

Information System). MEDIS implements CEN ENV 13 729 using X.509 public key certificates for authentication. For hierarchical role based access (RBAC) control MEDIS adopted XML by decomposing policy engines in to components. MEDIS uses WSS4J (Web Service Security for Java). WSS4J provides encryption and digital signing of SOAP message between central server and clinical server. The initial version of MEDIS simplifies administration and show acceptable performance in small community. However, integration of standards is required to implement MEDIS at country level. There is another design utilizes 128-bit symmetric encryption key into a code 128 barcode in Salford model [25] for access control on EHRs.

4.4.1.2.5 Cryptography

Suliman R et al [109] proposed a security architecture to secure communication in e-health. This architecture works at different security levels of EHRs. Then, symmetric cryptography is used for communication security and asymmetric cryptography used for authentication. This novel architecture shows potential when experimented with different algorithms. There is another design approach for secure sharing of EHR system is proposed by Chen Q et al [99]. This approach utilizes Elliptic Curves Cryptography and the AES encryption standard. Secure login, secure transmission of EHRs and reasonable control for authorization are key factors for this system. Whereas, proposed design do not implemented in real scenario. Riva et al [94] also proposed PING architecture for autonomy of patients. The PING server creates valid roles ***RV*** and evaluates supplied credentials for actor's identity. The interaction of ***RV*** and ***PR*** (PING record) under patient's obligations is resolute for access to EHRs. Whereas, authorization and authentication achieved by cryptography. The proposed solution is implemented as PING protocol. The protocol is successfully running in WWW environment. Two tier architecture Danilatou [69] also proposed for privacy and security in biomedical clouds which provides access control by cryptography. Through cryptographic signature and public key the credentials are verified to access specific EHR. The proposed solution use Key Note Trust Management System to show the process. The example shows that access control in de-centralized management can be achieve by cryptographic credentials.

4.4.1.2.6 Anonymization

Rashid H R et al [112] proposed conceptual framework with prototype to secure heterogeneous EHRs. This framework utilizes K-anonymization for de-identifying electronic health records. The release of k-anonymity conform to privacy policy generalize the release of EHRs. The proposed framework is evaluated with real dataset. Initial experiments results high accuracy for various attributes with flexible anonymization that guarantee privacy. Similarly, Song J et al [95] proposed security model for HIS of u-healthcare services. This model utilizes the anonymization and depersonalization to generalize the EHRs. Furthermore, for secure authentication technology like XML signature and certificate are used. The proposed HIS provide privacy protection, authentication and access control services for u-healthcare systems. In contrast, Xiao L et al [96] proposed data scheme based on link-anonymised with security model. This scheme enforces security and privacy policy in distributed EHRs. This scheme provided secure communication by using YellowPagesAgent . These agents are used to store private key and build trust relationship before communication. In this way, the proposed solution promises authorized access.

4.4.1.2.7 Web Services

Bohem O et al [64] proposed infrastructure for eCR system. This infrastructure gives federated authentication and authorization by using web services security technology. The system specifications are investigated from a case study. System shows validity for privacy principles in blueprint document. For access control of data for 24/7, a security method with 'Honest Broker' mechanism has been proposed by Boyd A. et al [97]. In this method MCRC (Michigan Clinical Research Collaboratory) is responsible for security. To encode and send messages, Web services are used for exchange of information. XML messages are wrapped in SOAP and HL7 standards followed for EHRs transmission. Authentications between systems are done with dual authentication option by using SSL handshake. Weaver C A et al [34] also proposed trusted secure networked based framework for HIS portal. This framework uses authenticated web services for access control. AES encryption along 256-bits key is used to secure network transmission. The proposed framework is validated in healthcare environment. The validation shows that HIPAA's security necessities will require essential changes and required trust level can control authorization.

4.4.1.2.8 Establish Trust

A hierarchical model [116] proposed for cross domain communication of healthcare units. According to this model node of hierarchy contains valid certificates in order to establish trust. In this architecture hop chain of trust is created among parent and child nodes. The established trust path of requester entity and final destination node (FDN) is validated for response. Moreover, symmetric algorithm and session tickets are used for encrypted data exchange. The qualitative comparisons of different Healthcare Units (HUS), each unit can retain its security policy while exchanging EHRs. Al-Zaharani S et al also proposed a architecture [23] to ensure and mange access control for de-centralized EHRs. This architecture used interactive trust negotiation method for authentication. This method utilizes cryptographic credentials to authenticate the stored EHRs. The proposed architecture is validated in simulated environment. 25 out of 275 were requests for access control. However, 17 requests were successful and 8 claims false requests. Furthermore, powerful interlopers were successful in braking the encryption security and firewall.

4.4.1.2.9 Authentication Certificate

Bhattacharya J et al [98] suggested middleware architecture as privacy broker. This architecture is EPAL based, capturing privacy policies. The privacy broker system capture user requirements and issue certificate for authentication of EHRs. This architecture force privacy policies in HIS to avoid privacy violation. In another study, Sheppard N P et al [35] proposed model by utilizing digital rights management to protect EHRs. In this model, work patterns are identified to issue license for authenticated roles.

4.4.1.2.10 Firewall

Maji K A et al [86] design a four tier architecture for web-bases telemedicine system iMedik. The proposed architecture is complied with HIPPA regulations. The three layers of architecture Database Layer, Business Layer and Presentation Layer are protected by firewall called Private Layer. This Private Layer protects EHRs from hacker attack. The proposed architecture is design to secure e-health application from hackers with dynamic availability. Meanwhile, a framework to empower patients with secure exchange of information via internet suggested by Ueckert F et al [110]. In this framework online EHRs divided in two different logical databases. First database contain patient personal information (name, address etc) and second databases have EHRs. Both databases are averted to use network by two level architectural firewall. However, network transport is protected by SSL-encryption. The proposed framework empowers patient to control their EHRs. This framework also increases information exchange among professional by EHR.

4.4.1.2.11 Ontology Based Technology

To make the usage of EHRs proficient and secure an ontology based technique proposed in [75]. In this technique Generic Medical Record Ontology (GMRO) is subdivided to make each EHR specific. This hierarchical ontology structure facilitates in access control. The technique proposed that ontology technology is significant for efficient and secure use of global EHRs.

4.4.1.2.12 Patients Define Access Conditions

O'Keefe M C et al [62] proposed an eConsent model to control access of de-centralized EHRs by patient's consent. This e-consent model makes e-consent objects (eCos) with the help of patient's formally stated consents and access control conditions. These eCos holds one or more rules which defines the user's access conditions under specific circumstances. The eConsent demonstrator successfully implemented, eConsent system proved secure sharing of EHRs under patient's wishes.

4.4.1.2.13 Behavior Based Access Control

Yarmand H M et al [34] evaluated behavioral based dynamic model for de-centralized EHRs. These novel model 'Context Aware System' (CAS) methods are used to model the behaviors of users. These CAS methods are used to define the logical constraints for access control. The proposed model is evaluated in real word case study. This model proposed flexible access control decisions according to behavior of users.

4.4.1.2.14 Situation Base Access Controls

A Situation Based Access Control model (SitBAC) by Peleg et al [83] balances the privacy and quality of EHRs, together. This model proposed a pattern having Patient, EHR, Access Task, Data Requester, Legal Authorization and Response with properties and their relations in Situation Schema. In this schema the access scenarios revolves around tasks as compare to roles.

The request to access EHRs is based on the situations. The proposed model control access according to situation within specific scenario.

4.4.1.2.15 Emergency Access Token

Huda N M et al [79] extend the architecture for EHRs. Their extension includes mechanism to access the EHRs in emergency situations. Emergency Access Token (EAT) is uses for this purpose. It checks the request value with emergency access and compare it with local EAT value and allow the access. The proposed mechanism is demonstrated which successfully entertain requests in emergency situations.

4.4.1.3 Implementation

4.4.1.3.1 Anonymization

A de-centralized framework [114] have de-identified attributes of EMRs by anonymzing. This anonymization model converted the original dataset 'D' into a transformed dataset 'Dthat' which generalized EHR. This conversion averted identity disclosure, guarantee privacy on retrieval of EHRs. Another, approach of Demuynck [70] for system which controls central repository of EHRs with strong access regulations by patient's themselves. To provide security, EHR belonging to same patients were made unlinkable through hash function H and a secret input skp(i). Only the authorized doctors; who were in possession of this secret input, could link and access those EHRs. The evaluation of this approach shows that both patients and doctors have concerns about privacy. The patients restrict access control by updating private key skp(i). However, except registration all communication of doctors with central authority is anonymous. Furthermore, Li R-Z et al [90] proposed a method for unlinkability between the patient and the EHRs in cloud computing platform. In order to restrict the disclosure of patient identity, EHRs are made anonymous. The method proved that cloud computing platform with unlinkability mechanism can be for secure exchange of EHRs.

4.4.1.3.2 Encryption

Lee B W and Lee C D [36] proposed a scheme to provide encryption of centralized EHRs. The proposed scheme provides cryptographic key management mechanism to fulfill HIPPA regulations. For confidentiality of Electronic health records (EHRs) digital signature system and symmetric cryptosystem are used. Thus, proposed schemes guarantee EHRs privacy by ingrained cryptographic mechanism. Ion also proposed a scheme [108] for encryption and access control of EHRs to enhance security of clouds which is holding EHRs. In this scheme table with attributes and names of patients are encrypted with pseudorandom function. Then, access rights are controlled by 'where' clause. The demonstration of the proposed scheme is implemented to allow SQL queries on encrypted databases with ownerships rights. This scheme guarantees to provide control on access of EHRs.

4.4.1.3.3 XML

For integrating healthcare information, Chi H et al [67] implemented Role Based Access Control model (RBAC). To protect resources XACML an OASIS standard is used which define policy language. However, XML scheme is used for access decision. The proposed model experiments in real world case scenario, model can be extended while protecting the computing environment. In another study, to accurately translate basic laws principles related to security safeguard and control Finance et al [73] proposed a XML based access control model to capture the patient consent according to need-to-know principles. Similarly, for implementation of EHRs privacy within the layered components of EHR framework XML security is used. XML based EHRs using XML security are created on local machine by owner's certificate and then server permits access control on selected EHRs to authorized users. Experimental results shows reduction of security cost.

4.4.1.3.4 Exponential and Laplace Noise

Bhasker et al has proposed a method [63] for discovering the k most frequent patterns for guaranteed privacy. In this algorithm he implemented exponential and Laplace noise-addition mechanisms which guarantee EHRs privacy during exchange of EHRs.

4.4.1.3.5 Task Independent Technique

To protect data privacy and data utility, Poovammal [85] implemented a task independent technique. To aid this principle, this technique transformed the sensitive attributes of EHRs into numerical value. This made both data private and utilizable.

4.4.1.3.6 Quantified Risk-adaptive Approach

A practical model has been proposed by considering realities to protect privacy of patients in HIS [115]. This model used quantified Risk-adaptive approach which periodically maintained risk score on each access of EHR. This risk score then used to determine the access control for next time.

4.4.1.3.7 Decision Learning Tree

Lindell [58] provided security in the process of data mining. Before sharing EHRs, a decision learning tree with ID3 algorithmic based model, identified and protected that information which could disclose patient's identity.

4.4.1.3.8 Secure-Dedup Algorithm

Malin B in 2010 develop Secure-Dedup algorithm [111] to produce k-unlinkable of EHRs for STRANON protocol. This algorithm is develop to fulfill the requirements of data protection model. It helps healthcare enterprises to restrict trail re-identification without enlightening identity. The proposed protocol is empirically evaluated in real world hospital scenario. However, results reveal that healthcare sector can share significant amount of EHRs without infringing anonymity.

4.4.1.3.9 Bipolar Multiple Number Base Technique

Chao M-H et al proposed [61] "bipolar multiple number base" technique for confidential exchange of EHRs over the Internet. This technique hide the EHRs into 'mark image' (usually mark of hospital). The digital signature is used to remove this mark image for EHRs authentication. The technique is validated through experiments. The results show that EHRs can be hided in image without any harm. As, comparison this technique is more conventional then traditional encryption/decryption except frequent change.

4.4.1.3.10 Pseudonymity

An e-prescription system [38] is proposed to protect patient privacy in drug prescription buy utilizing smart cards. The identity of patients and doctors are unlink able by pseudonymity. This unlink-ability hides patient identification.

4.4.1.3.11Context Specific Scheme

Deng M et al [37] proposed context-specific scheme for 'Federated Electronic Healthcare'(FEH). This scheme proposed identity management framework based on cryptographic algorithm. Hence, identities are issues according to the specific context for controlled authorization. The existing scheme in FEH is improved by introducing algorithm. This improvement do not privileged the attacker to correlate EHRs on large scale.

4.4.1.3.12Key Management Scheme

Cheng T-L [33] proposed a key management scheme to make mobile agents more secure and efficient access control. This key management scheme is based on Lagrange interpolation formula and hierarchal management structure which make efficient access control on EHRs by secure mobile agents.

4.4.1.3.13 Data and Data Structure Modification

A model based data analyses also proposed by Brumen [66]. We propose a solution for outsourced model-based data analyses. In this model data and data structure of EHRs are modified. Due to this modification privacy of EHRs are maintained during analysis.

4.4.2 Individual Participation and Control

The Individual Participation and Control principle was found to be the second most research principle with a total of 18 studies falling under this category. Most the studies in this category were related to SA&D (73%), IMPL (20%) and RE (6.6%).

4.4.2.1 Requirement Engineering

4.4.2.1.1 Data Ownership

The *data requester* of the security and privacy patterns proposed by Compagna [68] can only get the EHR if the *data owner* has given his consent to the *data provider*. Whereas, technique of Jenson et al [81] proposed that patients have rights to review and update data.

4.4.2.2 Software Architecture and Design

4.4.2.2.1 Patient Defined Access Control Rules

In Strawman design of Agrawal [20] each record is stored in table against a specific *'purpose'* attribute defined by the patient. These purposes are then combined in *'Authorization Tables'*, used for access control. This strategy gives complete ownership and control to patients at their own EHR. Ferrari [72] also suggested that when *access control rules (ACR)* are defined for EHRs are defined (See Principle: Security Safeguards and Controls), patients take part in defining these rules for access control. Similarly, the proposed e-consent model of O'Keefe M C et al [62] captures the consent of patient's before defining rules of access control mechanism. In this way, this model also provides a right to patients of whom and how their EHRs can be used. However, self determination model of Hass et al [77] allow patients to express, view, and alter privacy policy by policy management. Thus, proposed model facilitates patient to decide to use and exposé of their EHR for access control.

4.4.2.2.2 Pseudonymization

The Authentication layer of PIPE methodology [88] *(see principle security safeguards and control)* have three roles Affiliated (Relative), Data owner (Patient) and Authorized (HCP). Data owner have full control to authenticate other roles this layer. However, pseudonyms are used to encrypt and re-link the id of patients. In another study, PIPE architecture of Riedl et al [89] also guarantee privacy under strict control of patients. The patient control to her EHRs is achieved by layered structure of architecture. However, patient identification is hide by pseudonymization. The architecture of Alhaqbani et al [21] for linking EHRs in a way that gives patients control over their own EHRs . Indirect pseudonym identifiers are used for this purpose.

4.4.2.2.3 Access Management

Framework of Ueckert *(see principle security safeguard and control)* [110] also empowers patient to control their own EHRs as responsible citizens. Through access management patients himself define which parts of EHRs are accessible especially in emergency situations.

4.4.2.2.4 ChiperMe Technology

In ChiperMe architecture of Hensen I [32] store EHRs in separate boxes. The id of separate EHRs deposit boxes are in control of patients therefore, patients are fully empowered at their records.

4.4.2.2.5 I-Keys Hardware

Dalley A et al [29] model allow patients to withdraw and award access rights to GP by using I-Keys hardware. As a result the EHRs can be access at anytime and at any location on patient's choice.

4.4.2.2.6 Cryptographic Credentials

Two tier architecture of Danilatou [69] empower patients to issue access credential for users. The proposed solution guarantees access control in de-centralized management by cryptographic credentials under patient control.

4.4.2.2.7 Digital Rights Management

The model proposed by Sheppard N P et al [35] by utilizing Digital Rights Management (DRM). DRM model also consist Consent Management System (CSM) which records patient consent in *'Consent Objects'*. These objects are created by patients themselves. EHRs are released on the basis of these consents defined by the patients themselves.

4.4.2.3 Implementation

4.4.2.3.1 Anonymization

An approach by Demuynck [70] *(See Principle: Security Safeguards and Controls)* gives right to the patients to change the secret key and further reports to registrar by it. Although, patients could anonymously visit their doctors but the patient must give her secret key to him occasionally. The evaluation results that patient have to update its key after addition or alter of new records. As, to build trust with doctor patient have to share his new secret key.

4.4.2.3.2 Encryption

The proposed scheme of Lee [36] also support patient control. The use of key to decrypt the EHR is under control of patient. Patient has rights to control access of EHRs by using secret key for decryption. Thus, rights of patients are controlled by key usage of encrypted records. However, proposed scheme of Ion [108] for encryption and access control of EHRs in cloud computing environment. For ownerships rights the 'WHERE' clause of query is based on obligations of patients. This clause is checked before catering the requested query. Then access to EHRs allowed on the basis of this clause which is defined by patients themselves. Demonstration of this scheme is implemented to allow SQL queries on encrypted databases with ownerships rights.

4.4.2.3.3 Bipolar Multiple Number Base Technique

A secure data-hiding technique [61] by utilizing the bipolar multiple- base conversion provides patient privacy by hiding EHRs in mark image. Patients have the ownership rights to control the access of their EHRs. The mark is decrypted on the permission of patients to authorized users.

4.4.3 Use Limitation

A total of 14 studies were mapped under the Use Limitation principle. The studies under this principle were related to Requirement Engineering 14%, Software Architecture and Design 42%, and Implementation 42%.

4.4.3.1 Requirement Engineering

4.4.3.1.1 Assigned Duties

The data controller *(see principle: Accountability and Oversight)* grants limited use of data by allowing access on assigned duties in the security and privacy patterns proposed by Compagna [68].

4.4.3.1.2 Prioritize roles

Methodology of Braux et al [60] proposed to give statement level coverage from legal requirements. This methodology also proposed that EHRs released according to the need of request (e.g emergency situations). These need of requests are assessed by prioritizing roles like (nurses, doctors etc).

4.4.3.2 Software Architecture and Design

4.4.3.2.1 Purpose Based Access

The Record Access Control (RAC) of Strawman design [20] is responsible to limit the use of data. Every query to access EHRs is checked against record's *'purpose'* with information already specified in *Privacy Authorization Table.* After accessing purpose of the request the EHR is release, limits the use of EHR.

4.4.3.2.2 Smart Card Technology

A methodology using smart card technology [65] divided the entities in logical fragments (horizontally and vertically) in order to access need. There is a tag for each fragment with its cardinality. These fragments with need are allocated to the smart card. When this need match with information system record EHR is release for use.

4.4.3.2.3 Work Patterns

The model of Sheppard N P et al [35] disclose EHRs based upon *'need-to-Know' principle* by utilizing work patterns. Patients limit's the consent and disclosure of EHRs according to different work patterns. Furthermore, license to authenticate users, under patient's consent limits the use of EHRs.

4.4.3.2.4 Data Mining

To protect privacy in classification of EHRs a data mining technique is adopted by Khatri Ash et al [113]. The hybridization participating technique is used to separate local and global data. After this separation local and global rules are generated by horizontal and vertical mining technique. This technique helps in classification of EHRs, which limits its use when requested for access.

4.4.3.2.5 Ontology Based Technique

Ontology based technique *(see principle security safeguards and control)* of Hadzic [75] proposed that in order to access the EHRs, information can be presented in four different subontologies (Personal Information, Health Condition, Treatments and Appointments). This division further limits the access according to need resulting limited use of EHRs.

4.4.3.2.6 Encryption

Secure EHR system proposed [104] for patient data sharing in cross domain by preserving privacy. This system also allows minimum privilege delegation among specified roles of organizations PEKS-based access control. This minimum privilege on access control limits the disclosure of EHRs.

4.4.3.3 Implementation

4.4.3.3.1 Anonymization

Poovammal [85] implemented their task independent technique for two security principle. One is 'Security Safeguard and Control' and the other for this principle. This technique utilizes K-Anonymization and generalized by converting records in numerical values. These numerical values are then transformed in different ranges the EHRs. Only transformed values are published mapping values. These numerical values are then mapped according to need.

4.4.3.3.2 Quantified Risk-adaptive Approach

Wang [115] used quantified Risk-adaptive approach for use limitation of EHRs. The system using this approach calculated the risk score on 'need to know' principle. Every time access request recorded periodically and risk score is calculated for each request. This risk score with access control also utilizes to cater access request on need basis.

4.4.3.3.3 Enterprise Privacy Authorization Language (EPAL)

The Enterprise Privacy Authorization Language (EPAL) [98] is designed to capture privacy policies for the middleware architecture. In this language, SQL query is used to access request in which the regular SWL request is joined with *'Purpose'* attribute in where clause. The purpose attribute limits the use of EHRs.

4.4.3.3.4 Pseudonymity

The e-prescription system [38] proposed to protect patient privacy in drug prescription buy utilizing smart cards which hides patient identification by pseudonymity. This system also discloses necessary information to keep the confidentiality of EHRs.

4.4.3.3.5 XML

To accurately translate basic laws principles related to security safeguard and control Finance et al [73] proposed an access control model for XML [73] to capture the patient consent according to need-to-know principles.

Similarly, formal mapping of Croll et al [27] for patient privacy model. This model also map *need to know principle.* This principle results in limiting the use of sensitive medical information.

4.4.4 Accountability and Oversight

A total of 12 studies have been identified under the principle of Accountability and Oversight. Out of these, a total of 9 studies are categorized under Software Architecture and Design and 3 studies for Requirement Engineering Phase. However, no study reported this important principle for implementation phase.

4.4.4.1 Requirement Engineering

4.4.4.1.1 Oversight

Risk assessment methodology of Gritzalis S et al [30] proposed there should be oversight mechanism at communication channel to ensure security of EHRs. However, it is the responsibility of *data controller (See Principle: Security Safeguards and Controls)* in the security and privacy patterns proposed by Compagna [68] is to keep a check on the proper usage of EHRs.

4.4.4.1.2 Traceability

Technique of Jenson et al [81] to map legal requirements for sensitive health information provided traceability of EHRs usage through access log activities.

4.4.4.2 Software Architecture and Design

4.4.4.2.1 Auditing

Patients self-determination model of Hass et al [77] utilizes a symmetric fingerprint scheme to track disclosure of EHRs for third party. Tracking of records are used for auditing purpose. As compare to Hass model [77] to make the system of Straw man design [20] compliant, an audit trail is used in this system. This trail maintains all queries and the acceptance requests of users make to access EHRs in centralized database. The record of these queries and accepted requests for access are used for auditing. Similarly, in Choe's method [61] Central Access Control is able to perform auditing of Local Access Control (LAC) and Client applications by log transactions. However, the scheme of Grizalis et al [74] proposed that necessary auditing information should be maintained in logs by security agents in multiple policies domains. There is another analysis of Peyton L et al [84] suggested to introduced "Audit Service" architectural component in framework. This component also maintains logs of "Attribute Provider" for audit trials. Furthermore, context-specific scheme of Deng M et al [37] also proposed auditing mechanism for identity tracking. As, identity management protect the EHRS from attackers on large scale. Whereas, Blobel's approach [59] *(See Principle: Security Safeguards and Controls)* invocated audit policy control by identifying security policies which were related to message protection and security associations. This approach helps auditing of EHRs in large organizations. However, secure Client-Server architecture proposed by Lien C-C et al [100]. This architecture specifies the security requirements through roles. In this architecture component *AuditTrial* is responsible to record each access request, used for auditing.

4.4.4.2.2 Traceability

The proposed model of Song J et al [95] also dynamically trace EHRs. This tractability used for illegal disclosure and unauthorized access.

4.4.5 Data Integrity and Control

A total of 8 studies were mapped against the principle of Data Integrity and Quality. Out of these studies, 2 were placed under Requirement Engineering, 5 studies were categorized under Software Architecture and Design whereas, only one was categorized under Software Implementation.

4.4.5.1 Requirement Engineering

4.4.5.1.1 Secure Transmission

Jenson et al technique [81] ensure data control by checking for unauthorized manipulation during transmission.

4.4.5.1.2 Data Quality

The iTrust software requirements suggested by Massey [101] said to discard that EHRs which has been disabled for 7 years. This suggestion would improve the data quality of the database.

4.4.5.2 Software Architecture and Design

4.4.5.2.1 Signatures Schemes

In Choe's [61] method *(See Principle: Security Safeguards and Controls)* the request between Client Application and Central-Access control system is sent and received through XML digital signature to guarantee data integrity. The qualitative comparisons show each HCUs can retain its own security policy. Similarly, for content integrity XML signatures used to connect XML document with control policies in Gritzalis D et al [74] scheme. The proposed scheme supports integrity in multiple policy domains. Furthermore, The integrity of EHRs and data exchanged messages during interactions is guaranteed by signature schemes (i.e., HIDS, IBS) or message authentication code (i.e., HMAC) in EHR system of Sun et al [104].

4.4.5.2.2 Accurate EHRs

To address the principle of accuracy, Straw man design [20] also contained Data Accuracy Analyzer. The function of this analyzer was to update EHRs after or before entering the new EHR.

4.4.5.2.3 Asymmetric Fingerprinting Scheme

In Hass S et al [77] model *(See Principle: Security Safeguards and Controls)* integrity of data is maintained by sequence of tags for same records (delegation chain) during third party disclosure. To trace this disclosure is maintained by asymmetric fingerprinting scheme. However, model facilitates the completeness and correctness of EHRs on basis of system behavior.

4.4.5.3 Implementation

4.4.5.3.1 Cryptographic Key Management Scheme

In Lee B-W et al scheme [36] that integrity of EHRs are proposed by combination of cryptographic checksum and the symmetric cryptosystem.

4.4.6 Purpose Specification and Minimization

A total of 4 studies were mapped under the principle of Purpose Specification and Minimization. Out of the 4 studies 2 were related to Requirement Engineering, 1 for Software Architecture and Design and 1 related to Implementation.

4.4.6.1 Requirement Engineering

4.4.6.1.1 Minimum Information Collection

Risk assessment methodology of Gritzalis et al [30] also proposed that there is a minimum EHRs release according to the roles.

4.4.6.1.2 Compliance with Law

However, the methodology [87] proposed to comply software requirements with security policy. This methodology emphasized that purpose of EHRs collection should comply with law.

4.4.6.2 Software Architecture and Design

4.4.6.2.1 Patient Defined Purpose

Agrawal [20] proposed Straw man Design for 7 out of 9 principles for centralized EHRs. Except for ‘Remedies’ and ‘Openness and Transparency’, it catered all the other 7 principles. For this principle, this proposed to add an attribute ‘Purpose’ for each EHRs in authorization tables. This value of this attribute was specified by patients at the time of data collection. The reason for adding this attribute was to ensure the data entry on the basis of ‘purpose specification’.

4.4.6.3 Implementation

4.4.6.3.1 Purpose based collection

Mode also [82] *(See : principle safeguards and control)* defines an attribute of "purpose" in authorization policy for EHRs. This attribute is used to minimize collection of EHRs according to purpose.

4.4.7 Remedies

This principle also reported in 3 studies; which is related to Requirement Engineering and Software Implementation.

4.4.7.1 Requirement Engineering

4.4.7.1.1 Code of Ethics

Jenson and Massey also proposed that there should be a code of ethics (HIPPA) on illegal release of EHRs [81, 101].

4.4.7.2 Implementation

4.4.7.2.1 Copyrights

Framework of Bertino et al [31] for K-anonymity of patient data in order to safe patient privacy and copyrights. This framework used watermarking algorithm, provides copyrights protection in order to achieve rightful ownership.

4.4.8 Openness and Transparency

There are total of 3 studies were also categorized under the principle of *'Openness and Transparency'*. Two studies in this category are related to Software Architecture and Design and one of the study Implementation.

4.4.8.1 Software Architecture and Design

4.4.8.1.1 Public Key and Attributes Certificates

MEDIS architecture [106] *(see principle: security safeguards and control)* contain flag in distribution rules for EHRs by Public key and attributes certificates. These flags denotes whether patients are allowed to read architectural components. Through this type of permission EHRs flow is transparent to patients.

4.4.8.1.2 Access Management

However, the framework of Ueckert F et al [110] manages all users access to EHRs in same scenario. Due to access management of EHRs in same scenario, change in records is easy to understand for patients.

4.4.8.2 Implementation

4.4.8.2.1 Patient Understanding

In scheme [36] of Lee B W et al *(see principle: security safeguards and control)* patient understanding is developed about the notice in registration phase. In this notice patient is informed about users of EHRs. So, the schemes facilitate openness and transparency under compliance with HIPPA regulation.

4.4.9 Collection Limitation

Only 2 studies reported for this important principle; which is related to Software Architecture and Design and Implementation.

4.4.9.1 Software Architecture and Design

4.4.9.1.1 Consent and Disclosure Limitation

The founding principles of Strawman design *(see principle: Security Safeguard and Control)* [20] proposed for Hippocrate database. This principle covers collection limitation by consent and disclosure limitation. Privacy Constraint Validater (PCV) and Data Accuracy Analyzer (DAA) components of design are responsible to limit the consent and disclosure of EHRs.

4.4.9.2 Implementation

Privacy Model of Croll et al *(see principle: Security Safeguard and Control)* [27] also maps collection limitation principle of patients privacy requirements. This helps in specifying whether the relevant EHR (Electronic Health Record) is collected or not.

Privacy and Security Principles	Contribution by (Cluster name)	Proposed Solution	Study ID	Research Types	Frequency	SDLC Phase
Security safeguards and control	Authentication	Strong authentication by PKI-enables smart card.	[30]	Solution Proposal	1	Requirement Engineering
	Access Control	RBAC model for access control.	[30]	Solution Proposal	5	
		Check authorization level before access.	[81]	Solution Proposal		
		Role based access control to authorized roles.	[68]	Solution Proposal		
		Actor's hierarchies for access control.	[101]	Experience Paper		
		Define rules for access control	[60]	Solution Proposal		
	Secure data Transmission	Protect data transmission on communication channel.	[81]	Solution Proposal	2	
		Ensure data transmission security.	[87]	Solution Proposal		
Accountability and Oversight	Oversight	Oversight at communication channel.	[30]	Solution Proposal	2	Requirement Engineering
		Data controller for oversight.	[68]	Solution Proposal		
	Traceability	Traceability through Access log activities.	[81]	Solution Proposal	1	
Purpose Specification and	Minimum information release	Minimum information release according to roles.	[30]	Solution Proposal	1	Requirement Engineering

Minimization	Compliance of law	Purpose of data collection complies with law.	[87]	Solution Proposal	1	
Openness and Transparency	Transparent data collection	Client side protection measures by transparent data collection	[30]	Solution Proposal	1	Requirement Engineering
	Notify patients	Patients should be notifying about data storage.	[87]	Solution Proposal	1	
individual Participation and control	Data ownership	Data released on the evidence of data owner	[68]	Solution Proposal	1	Requirement Engineering
		Patient's rights to review and update data.	[87]	Solution Proposal	1	
Data integrity and quality		Check for unauthorized manipulation during transmission.	[81]	Solution Proposal	1	Requirement Engineering
		Automatically delete data after 6 years.	[101]	Experience paper	1	
Remedies	Code of ethics	An ethics code for privacy protection	[30]	Solution proposal	2	
		HIPPA law.	[101]	Experience paper		
Use Limitation	Assigned duties	Data access on assigned duties.	[68]	Solution Proposal	2	Requirement Engineering
	Rules prioritization	Priorities these rules to access records.	[60]	Solution Proposal		

Table 12 Distribution of Solution according to RE phase

Privacy and security Principles	Contribution by (Cluster name)	Proposed Solution	Study ID	Research Types	Frequency	SDLC Phase
Security safeguard and control	Smart card and biometrics for authentication and access control	Smart tokens for authorized access	[29]	Solution Proposal	4	Software Architecture and Design
		PKI, Smartcards and Biometrics	[71,	Solution		

		technology	78]	Proposal		
		smart card technology to reach the required security level.	[65]	Solution Proposal		
	Data encryption	Water marking for encryption	[77]	Solution Proposal	5	
		Data encryption with threshold scheme to avoid misuse	[75]	Solution Proposal		
		Homomorphic encryption scheme	[69]	Solution Proposal		
		Network-based encrypted objects through CipherMe extension of JavaScript functionality for data access and management.	[32]	Validation Research		
		Encryption of data on network for security and certificated issues for authentication	[22]	Solution Proposal		
		In this scheme table with attributes and names of patients are encrypted with pseudorandom function. Then, access rights are controlled by 'where' clause.	[108]	Solution Proposal		
		To guard the data from sneaking, this approach encrypted some data items	[20]	Solution Proposal		
		This approach utilizes Elliptic Curves	[99]	Solution Proposal		

		Cryptography and the AES encryption standard.				
	Pseudomization	Pseudonymlty for random ID	[77]	Solution Proposal	3	
		pseudonymous code (based on patient's identity) with encrypted communications	[93]	Solution Proposal		
		There is an encrypted link between identification and health pseudonyms. Through Pseudonymizatlon service, generalized EHRs are realized.	[88]	Validation Research		
	Established Trust	Established trust through global certificates for access control	[116]	Solution Proposal	3	
		This scheme provided secure communication by using YellowPagesAgent . These agents are used to store private key and build trust relationship before communication.	[96]	Solution Proposal		
		This architecture used interactive trust negotiation method for authentication.	[23]	Solution Proposal		
	Security agents	Security agents based on security service provider SSP(web services)	[74]	Solution Proposal	1	

		responsible for authentication and authorization.				
	Ontology based technology	Ontology based technology for authorized access control	[75]	Solution Proposal	1	
	Patient define access conditions	patient's formally stated consents and access control conditions	[62]	Solution Proposal	4	
		The interaction of ***RV*** and ***PR*** (PING record) under patients obligations are resolute for access to EHRs	[94]	Solution Proposal		
		The model proposed workflow based access control under strict obligations of patients.	[35]	Solution Proposal		
		This architecture specify the security requirements through rules.	[100]	Solution Proposal		
	Anonymization	The model utilizes the anonymization and depersonalization to generalize the EHRs.	[95]	Solution Proposal	2	
		The release of k-anonymity conform with privacy policy generalize the release of EHRs.	[112]	Solution Proposal		
	Web services	MEDIS uses WSS4J (Web Service Security for Java). WSS4J provides encryption and	[106]	Validation Research	4	

		digital signing of SOAP message between central server and clinical server				
		To encode and send messages, Web services are used for exchange of information. XML messages are wrapped in SOAP and HL7 standards followed for EHRs transmission.	[97]	Validation Research		
		The framework uses authenticated web services for access control. AES encryption along 256-bits key are used to secure network transmission	[107]	Validation Research		
		This infrastructure gives federated authentication and authorization by using web services security technology.	[64]	Solution Proposal		
	Behavior based access control	The novel model 'Context Aware System'(CAS) methods are used to model the behaviors of users. These CAS methods are used to define the logical constraints for access control	[34]	Evaluation Research	1	

	Role Separation	separated structural roles using organizational entity-to-entity relationship to assign specific roles and duties to each role. This separation of roles helps in management of authorization.	[59]	Solution Proposal	5	
		Access Control Roles (ACRs) must be thoroughly analyzed before they are defined.	[72]	Solution Proposal		
		Role-Based Access Control (RBAC) security model. This helped him in assigning the proper roles to users and managed the legal authorization in the system.	[102]	Solution Proposal		
		This architecture proposed the authorization then authentication followed by role privileges	[76]	Solution Proposal		
		Access control is given according to roles of these workgroup.	[105]	Solution Proposal		

	Capture requirements	The architecture is EPAL based, capturing privacy policies. The privacy broker system capture user requirements and issue certificate for authentication of EHRs.	[98]	Solution Proposal	1	
	Access Token	Emergency Access Token (EAT) is uses for this purposes, checks the request value with emergency access.	[79]	Solution Proposal	1	
	Situation based access	In this schema the access scenarios revolves around tasks as compare to roles. The request to access EHRs are based on the situations.	[83]	Solution Proposal	1	
	Cryptography	symmetric cryptography is used for communication security and asymmetric cryptography used for authentication.	[109]	Solution Proposal	1	
	Firewall	databases are averted to use network by two level architectural firewall.	[110]	Evaluation Research	2	
		The three layers of architecture Database Layer, Business Layer and Presentation Layer are protected by firewall called	[86]	Solution Proposal		

		Private Layer. This Private Layer protect EHRs from hacker attack.				
Accountability and Oversight	Biometrics	An a symmetric fingerprint scheme is used to track disclosure of EHRs for third party	[77]	Solution Proposal	1	
	Audit Trial	auditing information is maintained in logs by security agents in multiple policies domains.	[74]	Solution Proposal	4	
		AuditTrial are responsible to control accesses.	[100]	Solution Proposal		
		audit policy control by identifying security policies which were related to message protection and security associations. This approach helps auditing of EHRs in large organizations.	[59]	Solution Proposal		
		component maintains logs of "Attribute Provider" for audit trials	[84]	Evaluation Research		
	Tracability	tractability used for illegal disclosure and unauthorized access.	[95]	Solution Proposal	1	

Individual Participation and control	I-Key hardware	allow patients to withdraw and award access rights to GP by using I-Keys hardware	[29]	Validation Research	1	
	Pseudomyzation	Self determination model by utilizing Pseudomyzation	[77]	Solution Proposal	4	
		pseudonymization patient control at his/her own health records can be achieved.	[89]	Solution Proposal		
		the role of patient as owner by pseudonyms. Patient have full control at his records and grant access rights other roles of Authorization layer	[88]	Validation Research		
		the role of patient as owner by pseudonyms	[88]			
	cryptographic credentials	guarantee access control in de-centralized management by cryptographic credentials under patient control	[69]	Solution Proposal	2	
		PING server utilizes cryptography credentials to empowers patients	[94]	Solution Proposal		
	XML Signature	separate EHRs deposit in XML objects boxes are in control of patients	[32]	Validation Research	1	

	Capture Consent	consent of patient's before defining rules of access control mechanism.	[62]	Solution Proposal	1	
	Role bases Access Control Model	Server empowers patients by write access.	[110]	Solution Proposal	3	
		mange all users to access EHRs in same scenario. This type of management allow patients to tracks changes in his/her records.		Solution Proposal		
		They provided solution for this principle by introducing Patient Medical Record and Patient Information Service in their **framework** to facilitate patients in obtaining their EHRs, via internet	[102]	Solution Proposal		
Data Integrity and quality	Tracking	integrity of data is maintained by tracking sequence of tags for same records (delegation chain)	[77]	Solution Proposal	2	
		FDN is used for track the changes in EHRs	[116]	Solution Proposal		
	XML Signatures	For content integrity XML signatures used to connect XML document with control policies	[74]	Solution Proposal	1	
Use Limitation	Ontology based hierarchy	Hierarchical structure of ontology can be	[75]	Solution Proposal	2	

		further divided to limit the use of EHRs.				
		mange all users to access EHRs in same scenario. This type of management allows patients to tracks changes in his/her records.	[110]	Solution Proposal		
	Smart card Technology	using smart card technology (Bolchini) controlled the access of EHR on 'need-to-know' principle	[65]	Solution Proposal	1	
	Data Mining Technique	hybridization participating technique is used to separate local and global data	[113]	Solution Proposal	1	
Openness and Transparency		Attributes certificates of MEDIS architecture contain flags which allow patients to read architectural components.	[106]	Solution Proposal	2	
		Management allow patients to tracks changes in his/her records.	[110]	Solution Proposal		

Table 13 Distribution of Solutions According to SA&D

Privacy and Security Principles	Contribution by (Cluster name)	Proposed Solution	Study ID	Research Types	Frequency	SDLC Phase
Security Safeguard and control	Anonymization	A de-centralized framework have de-identified attributes of EMRs by anonymzing. This anonymization model converted the original dataset '**D**' into a transformed dataset 'Dthat' which generalized EHR.	[114]	Validation Research	3	Implementation
		An approach for a system which controls central repository of EHRs with strong access regulations by patient's themselves.	[70]	Solution Proposal		
		a method for unlinkability between the patient and the EHRs in cloud computing platform. In order to restrict the disclosure of patient identity, EHRs are made anonymous. The method proved that cloud computing platform with unlinkability mechanism can be for secure exchange of EHRs	[90]	Solution Proposal		

	Encryption	proposed a scheme to provide encryption of centralized EHRs. The proposed scheme provides cryptographic key management mechanism to fulfill HIPPA regulations.	[36]	Solution Proposal	2	
		A scheme for encryption and access control of EHRs to enhance security of clouds which is holding EHRs. In this scheme table with attributes and names of patients are encrypted with pseudorandom function.	[108]	Solution Proposal		
	XML	To protect resources XACML an OASIS standard is used which define policy language. However, XML scheme is used for access decision.	[67]	Solution Proposal	2	
		A XML based access control model to capture the patient consent according to need-to-know principles	[73]	Solution Proposal		
	Exponential and Laplace Noise	A method for discovering the k most frequent patterns for guaranteed privacy. In this	[63]	Evaluation Research	1	

		algorithm he implemented exponential and Laplace noise-addition mechanisms which guarantee EHRs privacy during exchange of EHRs				
	Task Independent Technique	To protect data privacy and data utility, implemented a task independent technique. To aid this principle, this technique transformed the sensitive attributes of EHRs into numerical value. This made both data private and utilizable.	[85]	Solution Proposal	1	
	Quantified Risk Adoptive Approach	A practical model has been proposed by considering realities to protect privacy of patients in HIS. This model used quantified Risk-adaptive approach which periodically maintained risk score on each access of EHR. This risk score then used to determine the access control for next time.	[115]	Evaluation Research	1	
	Decision Learning Tree	Security provided in the process of data mining. Before sharing	[58]	Evaluation Research	1	

		EHRs, a decision learning tree with ID3 algorithmic based model, identified and protected that information which could disclose patient's identity.				
	Secure Dedup Algorithm	Secure-Dedup algorithm is proposed to produce k-unlinkable of EHRs for STRANON protocol. This algorithm is develop to fulfill the requirements of data protection model. It helps healthcare enterprises to restrict trail re-identification without enlightening identity.	[111]	Evaluation Research	1	
	Bipolar Multiple Number Base Technique	Bipolar multiple number base" technique for confidential exchange of EHRs over the Internet. This technique hide the EHRs into 'mark image' (usually mark of hospital). The digital signature is used to remove this mark image for EHRs authentication.	[61]	Solution Proposal	1	

	Pseudonymity	An e-prescription system is proposed to protect patient privacy in drug prescription buy utilizing smart cards. The identity of patients and doctors are unlink able by pseudonymity. This unlink-ability hides patient identification.	[38]	Solution Proposal	1	
	Context Specific Scheme	Deng M et al proposed context-specific scheme for 'Federated Electronic Healthcare'(FEH). This scheme proposed identity management framework based on cryptographic algorithm. Hence, identities are issues according to the specific context for controlled authorization.	[37]	Solution Proposal	1	
	Key Management Scheme	Cheng T-L proposed a key management scheme to make mobile agents more secure and efficient access control. This key management scheme is based on Lagrange interpolation formula and hierarchal management structure which	[33]	Solution Proposal	1	

		make efficient access control on EHRs by secure mobile agents.				
	Data and Secure Structure Modification	A model based data analyses proposed a solution for outsourced model-based data analyses. In this model data and data structure of EHRs are modified. Due to this modification privacy of EHRs are maintained during analysis.	[66]	Solution Proposal	1	
Use Limitation	XML	an access control model for XML to capture the patient consent according to need-to-know principles	[73]	Solution Proposal	2	
		This model also map *need to know principle*. This principle results in limiting the use of sensitive medical information.	[27]	Solution Proposal		

	Pseudonymity	The e-prescription system proposed to protect patient privacy in drug prescription buy utilizing smart cards which hides patient identification by pseudonymity. This system also discloses necessary information to keep the confidentiality of EHRs.	[38]	Solution Proposal	1	
	EPAL	The Enterprise Privacy Authorization Language (EPAL) [98] is designed to capture privacy policies for the middleware architecture. In this language, SQL query is used to access request in which the regular SWL request is joined with *'Purpose'* attribute in where clause. The purpose attribute limits the use of EHRs.	[98]	Solution Proposal	1	
	Quantified Risk Adoptive approach	A quantified Risk-adaptive approach for use limitation of EHRs. The system using this approach calculated the risk score on 'need to know'	[115]	Evaluation Research	1	

		principle. Every time access request recorded periodically and risk score is calculated for each request. This risk score with access control also utilizes to cater access request on need basis.				
	Anonymization	This technique utilizes K-Anonymization and generalized by converting records in numerical values. These numerical values are then transformed in different ranges the EHRs. Only transformed values are published mapping values. These numerical values are then mapped according to need.	[85]	Solution Proposal	1	
Data integrity and quality	Cryptography Key Management Scheme	The integrity of EHRs are proposed by combination of cryptographic checksum and the symmetric cryptosystem.	[36]	Solution Proposal		
Purpose specification and minimization	Purpose based collection	Mode defines an attribute of "purpose" in authorization policy for EHRs.	[82]	Solution Proposal	1	

		This attribute is used to minimize collection of EHRs according to purpose.				
Openness and transparency	Patient Understanding	In scheme of Lee B W et al patient understanding is developed about the notice in registration phase. In this notice patient is informed about users of EHRs. So, the schemes facilitate openness and transparency under compliance with HIPPA regulation.	[36]	Solution Proposal	1	
Remedies	Copyrights	Framework for K-anonymity of patient data in order to safe patient privacy and copyrights. This framework used watermarking algorithm, provides copyrights protection in order to achieve rightful ownership.	[31]	Solution Proposal	1	
individual Participation and Control	Anonymization	An approach gives right to the patients to change the secret key and further reports to registrar by it. Although, patients could	[70]	Solution Proposal	1	

		anonymously visit their doctors but the patient must give her secret key to him occasionally.				
	Encryption	The use of key to decrypt the EHR is under control of patient. Patient has rights to control access of EHRs by using secret key for decryption. Thus, rights of patients are controlled by key usage of encrypted records.	[36]	Solution Proposal	2	
		Scheme for encryption and access control of EHRs in cloud computing environment. For ownerships rights the 'WHERE' clause of query is based on obligations of patients.	[108]	Solution Proposal		
	Bipolar number based scheme	A secure data-hiding technique by utilizing the bipolar multiple-base conversion provides patient privacy by hiding EHRs in mark image. Patients have the ownership rights to control the access of their EHRs. The mark is decrypted on the permission of patients to	[61]	Solution Proposal	1	

5 Validity Threats

This chapter reports how validity threats are identified and minimized during this systematic literature review. These threats are divided into four main categories Investigation Bias, Publication Bias, Primary Studies Selection Threats and finally Data Extraction Threats.

5.1 Investigation Bias

As this SLR due to degree requirement conducted by an individual author, there is a probability of more validity threats (judgment, bias, anchoring etc) as compare to those SLRs conducted by more than one researcher. This type of biasness effect primary study selection and the data extraction phase, which will be explained in subsequent sections. However, different steps and results of SLR phases are thoroughly checked by other SLRs experts. Abstract reading for initial selection and full reading for data extraction phases was conducted several times to minimize the investigation bias. However, it can't guarantee to solve validity threats but to some extent reduce it.

5.2 Publication Bias

The publication bias is addressed to minimized the selection of negative research outcomes and maximized the outcomes of positive publications [11]. First this bias is addressed by phrasing general question (RQ1) to cover all phases of SLC. We do not rely any specific phase of SDLC. Secondly, choose patient privacy and security principles of Markle foundation [10] to cover all aspects of patient privacy. As, aforementioned principles cover all major patient privacy laws implemented in Europe, Australia and USA. Due to this general nature of SLR, studies covering patient data privacy and security for any phase of software system development are part of our review. Furthermore, this research is restricted to four major databases. These databases are included to the importance and relevance of the concerned area. However, reliability of the retrieved studies from these databases is checked against already known set of papers (mapping study) [53].

5.3 Threats to Identification of Primary Studies

As mentioned earlier, it was a challenge to generate a valid search string which retrieves all studies patient data privacy. Huge numbers of terms have been used for Electronic Health

Records (EHRs) by literature. Therefore, keywords of 58 selected papers of mapping study [53] are used to generate a valid and a reliable string. For instance ((Health Records" OR" Patient Information" OR "Medical Records" OR" Electronic Healthcare Records" OR" Electronic Health Records" OR" Electronic Medical Records" OR" Patient Data" OR" Medical Data" OR" Protected Health Information") AND (Privacy OR Security)) is the search string used to minimize primary studies selection threats from databases. However, there is s limitation to select those studies which do not allow full text download.

5.4 Threats to Data extraction and Result consistency

A reviewed protocol is necessary to have a more effective and efficient SLR [11]. As mention in chapter 2, selections of studies were based on inclusion and exclusion criteria for consistent data extraction. The protocol and its criteria were based upon the opinion of SLR's experts. A data extraction form was designed and evaluated in pilot extraction of 10 papers randomly chosen from databases. Therefore, piloting consistent the results in data extraction phase. To increase the reliability and to minimize the data extraction bias, 77 independent studies were validated. Out of 77 studies 35% were validated by study's authors and 65% were validated by fellow experts. Consequently, iterative improvement in data extraction increases the result reliability and consistency.

6 Discussion and Conclusion

This chapter briefly discusses major findings of this thesis. The results suggested need of major research in patient data privacy and security solutions for software system development. Finally, highlights some areas for future work.

6.1 DISCUSSION

This review presents a catalogue with state of-the-art practical solutions to secure EHRs in HIS. While performing this review, we found that environment for de-centralized EHRs received more attention than centralized EHRs. Similarly, the software architecture and design gain much attention as compared to implementation and requirement engineering phases. No study reported on testing and maintenance phases of SDLC. This may be cause of concern for both the researchers and the practitioners as most defects in requirements are introduced at requirements phase and one of the most effective ways to give assurance of compliance with desired standards and requirements is Testing and Quality Assurance.

Furthermore, we have also found trend of using cloud computing technology for large storage and quick access of health records in distributed environment [69, 108]. We have observed that for 'Openness and Transparency' principle no general policy was adopted, individuals were informed who can access and use their data but did not notify where their records resides and for what purpose it can be used.

For ' Purpose Specification and Minimization'; purpose was minimized for use of data but did not specified at time of collection and change in purpose. In privacy principle 'Individual Participation and Control' patients have right to control access of their data but did not provided rights that health data communicated with them in understandable form occasionally, patients can know reasons on denial of request and challenge when amendments were made in their data. It is shown that patients had some rights to control their information while, no rights have been allocated for healthcare providers to control their information. Similarly, 'Data integrity and control' was provided by only completeness and accuracy of data while, less focus paid to update data. Authorization, access control and encryption of data were reported for 'security safeguards and control' whereas, less focus paid when data had to be accessed in emergency situations and data security on networks. Legal remedies were reported but no study addressed financial remedy on privacy violation.

We believe that this review provides practical solutions to implement patient data privacy and security for EHRs in HIS. However, there is a great need to elicit requirements for centralized EHRs. Research community should pay special attention towards some important privacy principles like 'Collection Limitation', 'Use Limitation', 'Data Integrity and Quality', 'Openness and Transparency' and 'Remedies'. Research community should also focus on detail specifications of aforementioned principles. Similarly, centralized EHRs need more practical solutions. Furthermore, same privacy policy should be adopted for exchange of EHRs. Finally, authorization and access control under 'Security Safeguards and Control' should also incorporate mechanism to access EHRs in emergency situations.

6.2 CONCLUSION

This systematic review shows evidence for patient data privacy and security to develop HIS. However, the literature gives broad coverage of all nine predefined principles. Real threats to patient data privacy and security can be eliminated by focusing research at detail specifications of each nine principles. 'Security safeguards and controls' and ' Individual participation and control' have got real attention as compared to other privacy principles. Similarly, large number of studies reported for software architecture and design phase, while, no study proposed solution for software testing and maintenance phase. Meanwhile, research community is more focused to answer privacy and security for de-centralized EMRs as compared to centralized EMRs.

6.3 Future Work

This systematic Review highlights some important research areas which need serious attention to fulfill all aspects of patient data privacy and security.

- Important privacy principles like collection limitation, openness and transparency, purpose specification and minimization and remedies need serious attention from research community.
- Latter phases of SDLC needed practical solutions to answer patient data privacy and security in order to improve the quality and maintenance of system developed for healthcare sector.
- There is also a great need to observe usage of EHRs in real scenarios. As, to make implementation of privacy laws more practical for healthcare sectors.

- And finally, there is also a great need to evaluate the impact of these privacy and security laws on patient safety.

7 Appendix 1

PROTOCOL

Systematic Literature Review of Patient Data Privacy and Security for Software System Development

Isma Masood[1]
Saad Zafar[2]

External Review

Barbara Kitchnham

Dr. Mehmood Niazi

Kelee University UK

Protocol Version 4

April, 201

[1] Department of Computer Sciences & Software Engineering, Faculty of Basic and Applied Sciences International Islamic University Islamabad, Pakistan.

[2] Faculty of Computing, Riphah International University Islamabad, Pakistan.

Protocol: Systematic Literature Review of Patient Data Privacy and Security for Software System Development

Preamble

Patient data privacy and security in Electronic Health Records (EHR) are facing lots of threats due to its growing importance. Therefore, research literature brings number of solutions from last ten years to answer patient data privacy and security. As the field of software Engineering shows less empirical evidence therefore, the proposed study bring together published work on solutions of patient data privacy and security for software system development by following systematic literature review guidelines[118] for the first time. The aim is to summarize research studies related to our research questions in fair, rigorous and auditable manner.

As recommended by [118] we consider the broad repercussion of our research. In this study, a Systematic Literature Review will be conducted on solutions of patient data privacy and security proposed for software system development. This study has a broad justification because this study will helps software engineers to understand what different contributions and solutions have been proposed in literature for patient data privacy and security, which can lead them to a better set of core elements as well as practical approaches for software system development. A synthesis of the empirical literature for software system development to answer patient data privacy and security will be able to identify gaps and current state of practice.

Background

Literature defines the individual's privacy as one's personal space free from interference of other people and organizations which are fundamental rights of any individual that needs to be protected [119]. Privacy considers a key requirement in patient physician relationship. Shared information of patients is use for correct diagnosis and treatment and to avoid adverse drug interactions. Sometimes patients may refuse to disclose important information in cases of health problems such as psychiatric behavior and HIV as disclosure, which may lead to social stigma and discrimination. Patients health records are not only used for diagnosis and treatment but also used for many other purposes, for instance to improve efficiency within the healthcare system, drive public policy development and administration, conduct of medical research, payer organizations to justify payment of services rendered etc [120]. Nowadays, patient's personal medical information is entered, processed, accessed and transmitted electronically. New challenges have threatened the individual's privacy rights in this flow of information of healthcare sector. Maintaining privacy and security of patient's personal information becomes more challenging. Electronically shared information within healthcare sector and other organizations are receiving threats to patient data privacy and security, in a survey[120] these threats are categorized as organizational and systematic threats. In another study [121] information security threats in network system of medical organizations are categorized in according to events i.e. :-natural events, external events and internal events. Threats to patient data privacy and security becomes a major cause of inaccuracies and improper disclosure of information which threaten individual's personal life and financial well being[122]. Therefore, many laws and policies in different countries have been implemented to protect patient data privacy and security especially for Electronic Health Records (EHR) [123].

Software system developed for healthcare sector counter some principles of these laws and policies still there is a great need for effective integration of privacy and security principles in software systems.

In order to obtain a real view of existing research we want to conduct a systematic literature review (SLR) on solutions of patient data privacy and security for software system development. No SLR in literature has been reported earlier in this context.

We therefore, construct two research questions to gain a broad view of the research on patient data privacy and security for software system development.

Research Questions

Our initial research question closely match the type suggested by [118] where the emphasis is accessing how technology is in/affect software engineering. We have closely followed the PICOC (Population, Intervention, Comparison, Outcome, and Context) criteria to frame the research questions suggested in [4]. However, we have excluded Comparison from 1st research question depending upon the nature of systematic study. To assess the current state of practice in software system development for patient data privacy and security, we have formulated the following two research questions:-

RQ.1. ***which solutions of patient data privacy and security have been proposed for software system development?***

Population: Empirical Literature for solutions of patient data privacy and security.

Intervention: Solutions to answer privacy principles – Stages of software system development.

Comparison: No.

Outcome: Solutions for patient data privacy and security propose for software system development.

Context: All type of empirical and theoretical studies.

RQ.2. ***Can we categorize these solutions using the Markle Foundation's Common Framework?***

Population: List of empirical solutions on patient data privacy and security for software development.

Intervention: Solutions to answer privacy principles – Stages of software system development.

Comparison: on the basis of privacy policy principles.

Outcome: Solutions for patient data privacy and security propose for software system development to answer privacy policy principles.

Context: All type of empirical and theoretical studies.

Search for Primary Studies

The traditional criteria defined in guidelines of Babara Kitchamn [4] to drive search terms for making search strings was unable to give desired results. Therefore, our search string terms have been derived from our mapping study [4].

3.1. Systematic Mapping

The goal of performing systematic mapping study is to justify this area for SLR, identify the quantity, type of research and results available within it and to show frequencies of publications by focusing trends [118, 124] for full mapping study see appendix B. The results of mapping study have been published in [53].

3.2. Search Strategy

The strategy used to construct search terms is as follows:

Derive relevant major terms and keywords from studies of systematic mapping study to access full breath of topic.

When database allows, use the Boolean OR to incorporate alternative derived terms.

When database allows, use the Boolean AND to link the major terms from population, intervention and outcome.

Results:

Following are the major terms derived on the basis of our systematic mapping study:-

((Health Records" OR" Patient Information" OR "Medical Records" OR" Electronic Healthcare Records" OR" Electronic Health Records" OR" Electronic Medical Records" OR" Patient Data" OR" Medical Data" OR" Protected Health Information") AND (Privacy OR Security))

3.3. Research Sources

Following research databases are used to access the primary studies:-

- IEEE Explore
- ACM Digital library:

- ScienceDirect
- SpringerLink

Scope: To avoid duplication and bias in reference validation, we have avoided to select the large number of sources and search conferences manually because different sources have different standards for reporting author names (some start with given names, some start with family names) and for referencing conference papers (sometimes the conference is used, sometimes the proceedings names are used). We also exclude books from our review of the literature.

3.4. Search Process Documentation

The search process involves two stages. Stage one: A primary search on the 'databases' listed in 3.2. Stage two: Secondary searches made as a result of identifying work in our primary search.

3.4.1. Primary Search documentation

We document our primary search as follows:-

3.4.1.1. Document: search terms

<table>
<tr><th>Date</th><th>Search String
TI=within title|AB=within Abstract| All= within ALL ,</th><th>Search Identifier</th><th>Comment</th></tr>
<tr><td rowspan="5">15th Sept,2011</td><td rowspan="5">"Abstract":((Health Records" OR " Patient Information" OR "Medical Records" OR " Electronic Healthcare Records" OR " Electronic Health Records" OR " Electronic Medical Records" OR " Patient Data" OR " Medical Data" OR " Protected Health</td><td>AB</td><td>RQ1-search</td></tr>
<tr><td colspan="2">1 separate searches (2000-2011)</td></tr>
<tr><td>IEEE</td><td>367</td></tr>
<tr><td><b>Total</b></td><td><b>367</b></td></tr>
<tr><td></td><td></td></tr>
</table>

	Information") AND (Privacy OR Security))		

Table 15 Look-up Table for RQ1

We have tailored our main search string according for each database.

Following are the look up table which shows that how string are search in databases against each research questions. We store as much information as possible about each paper by using tool Endnote.

I. CUT AND PASTE THE **SEARCH STRING** INTO DATABASE ACCORDING TO FIELDS.

3.4.1.2. **Document: Default Endnote fields are expanded to include the following:**

Author:

Year:

Title

Paper ID:

Journal/Conference:

Publisher:

Volume:

Issue:

Pages:

Date (of conference):

Researcher Name:

Date of Search:

Search String: Lookup Table Ref #

Inclusion Criteria1: The study belongs to theoretical or empirical studies?

Inclusion Criteria2: The study focuses on any stage of software system development for Patient data privacy and security?

Inclusion Criteria3: The study answers any privacy principle given in [9]?

Inclusion Criteria4: The study been published in between 2000-2011?

Exclusion Criteria1: If the study's proposed privacy and security solution support other data rather than patient's data?

Exclusion Criteria2: If the study supports other fields rather than software engineering?

Exclusion Criteria3: If the study supports other areas of software engineering rather than any development stage of software system?

++Quallty Criteria (score):

++Type of Study: (Validation Research/ Evaluation Research/ Solution Proposal/ Philosophical paper/ Opinion Paper/ Experience report):

Decision Based on (Title/Abstract/Introduction/Conclusion/Whole Paper):

+++repeated study:

***population:**

Decision Status (Include/Reject/Waiting for Full paper/Don't Know):

***Keywords:**

***Abstract:**

***Notes:**

URL:

~*Review Guidelines: (pdf file linked to Endnote for easy reference)

>>++Quality Assessment and Results form: (Embedded 'Quality Assessment and Accepted papers and Follow-up form')

***Name of Reference Database:**

KEY:

* = optional fields (dependent on type of study and available data);

~ = you need to make link from your stored file;

+ = only if paper passes exclusion criteria test;

++ = only if paper passes exclusion AND inclusion tests

+++ = only if paper passes exclusion AND inclusion tests AND quality criteria assessment made.

>> = embedded file

All other fields are compulsory

3. Study Selection Criteria

Following are some important points which we consider while including or excluding primary studies:-

3.1. Inclusion Criteria

a. The study belongs to theoretical or empirical studies?

b. The study focus on any stage of software system development for Patient data privacy and security?

c. The study answers any privacy principle given in [9]?

d. The study been published in between 2000-2011?

3.2. Exclusion Criteria

a) If the study's proposed privacy and security solution support other data rather than patient's data?

b) If the study supports other fields rather than software engineering?

c) If the study supports other areas of software engineering rather than any development stage of software system.

4. Study Selection Procedures

Following study selection criteria will be applied to select the primary studies.

4.1. Study Quality Assessment Criteria

To find actual evidence and relevance of selected studies quality assessment will be done [4]. Furthermore, quality assessment will also provide sound and reliable evidence of selected studies to answer the research questions. The quality assessment process will be validated of all selected studies by their authors or by fellow experts.

4.2. Inclusion and Exclusion Criteria

To minimize the bias effect while conducting review Inclusion and exclusion criteria will be applied (see section 4). During search process this criteria can be refined.

4.3. Primary Selection Process

During the primary selection process, screening of title abstract and keywords as filter 1 will be performed. After applying filter 1 studies will be added in final reference library however, record of the rejected studies will also be maintained.

4.4. Final Selection Process

In case, if study does not meet Inclusion criteria whole paper will be read, selection at this stage will be added in 'final reference library'. Any confusion about selection of studies will be discussed with supervisor.

5. Quality Assessment Checklists and procedures

This section presents the quality assessment checklists for both qualitative and quantitative studies.

5.1. Document: Quality Assessment Form

Forms which is shown in Table 21 and 22 were embedded in the 'Results and Quality Assessment Form' field in Endnote based on the quality check list provided by in (protocol of Mehwish Riaz). The form will be completed for ALL papers that have been added to 'Final reference library'. The quality assessment form lists and aggregates quality criteria. The objective is to provide a rough assessment to the quality of the paper before completing the accepted paper form. This assessment does not act as an exclusion criterion but guides elucidation. The score alone has little meaning; to understand the quality we need to look at the criteria and context of the assessment and cannot compare quality of different papers as based on the score alone.

Endnote has only one field that allows a file attachment. We will use this field for both this quality form AND the accepted papers form (explained in next section).

Sr no.	Questions	Answers Yes=1, No=0, Partially= 0.5

1	Are the research question(s) clearly stated for the studies?	Yes/No/Partially
2	Does the study build upon existing body of knowledge, i.e., does it explicitly discuss its contribution in the light of previous work?	Yes/No/Partially
3	Are the variables/metrics used in the study adequately measured and validated?	Yes/No/Partially
4	Are the metrics used in the study clearly defined?	Yes/No/Partially
5	Are all model construction methods/metric(s) derivation methods fully defined (tools and methods used)?	Yes/No/Partially
6	Are the metrics used in the study the most relevant ones for answering the research questions?	Yes/No/Partially
7	Are the data collection methods adequately described?	Yes/No/Partially
8	Are the statistical methods justified?	Yes/No/Partially
9	Is the purpose of the data analysis clear?	Yes/No/Partially
10	Are potential confounders adequately controlled in the analysis?	Yes/No/Partially
11	Are the negative findings presented?	Yes/No/Partially
12	Do the researchers discuss any problems with the validity/reliability of their results?	Yes/No/Partially
13	Is the study replicable?	Yes/No/Partially
14	Is the research design clearly presented?	Yes/No/Partially
15	Is the research design suitable for carrying out the study?	Yes/No/Partially
16	Are the findings credible?	Yes/No/Partially
17	Is the research process described thoroughly? Are the roadblocks, false steps described in a helpful way?	Yes/No/Partially
18	Are the links between the data, interpretation, and conclusions clear?	Yes/No/Partially
19	Is the reporting clear and coherent?	Yes/No/Partially
	Total Score	**Score**

Table 16 Quality Assessment Form for Quantitative Studies

Sr no.	**Questions**	**Answers** Yes=1, No=0,

		Partially= 0.5
1	Is the research design suitable for carrying out the study?	Yes/No/Partially
2	Does the study build upon existing body of knowledge, i.e., does it explicitly discuss its contribution in the light of previous work?	Yes/No/Partially
3	Does the study report clear, unambiguous findings based on evidence and argument?	Yes/No/Partially
4	Are the findings credible?	Yes/No/Partially
5	Is the research process described thoroughly? Are the roadblocks, false steps described in a helpful way?	Yes/No/Partially
6	Are the links between the data, interpretation, and conclusions clear?	Yes/No/Partially
7	Is the reporting clear and coherent?	Yes/No/Partially
8	Are the assumptions/theoretical perspectives/values that have shaped the form and output of the evaluation clear?	Yes/No/Partially
	Total Score	Score

Table 17 Quality Assessment Form for Qualitative Studies

6. Data Extraction Strategy

After selection and quality assessment of Primary studies, the data will be extracted. This section presents data extraction forms and data extraction procedure.

6.1. Document: Accepted papers/Follow-up Form

This form, shown in Table 23, will be embedded in the 'Results Quality Assessment' field in Endnote (with the Quality Assessment). This will ensures that all documentation relating to each paper is stored at one place.

Reviewer Name:	
Title of Paper:	
Paper ID:	
THE FQLLOWING REFER TO OUR RQs:	RECORDED IN PAPER

1	**(RQ1)** solution type (technique/methodology/process/model/approach/framework/ method/scheme)	
2	**(RQ1)** Software Development stage (Requirement lifecycle/design / implementation)	
3	**(RQ2)** privacy and security rules 1) Openness and Transparency 2) Purpose Specification and Minimization 3) Collection Limitation 4) Use Limitation 5) Individual Participation and Control 6) Data Integrity and Quality 7) Security Safeguards and Controls 8) Accountability and Oversight, 9) Remedies.	
4	Research Type	
5	Context: (Distributed network, E-Health, Internet bases etc)	
6	Other observation	
7	References found in paper (to follow up)	

Table 18 Generic Form for Results of Accepted Papers

Instructions:

If we fill in any of the fields 1 – 6,

a. Save soft copy of the Quality Assessment and Results form in Accepted papers folder – along with the full version of the paper.
b. Embed electronic copy of this form into Endnote Quality and Results Form field. If you fill in secondary sources field (7), make a copy of the file and save in Secondary Sources folder, for later follow up work.

6.2. Document: Secondary Search

This is similar to primary search documentation, other than no search string/lookup table will be used. Endnote is used in the same way to record the references as for primary studies. The one exception is that for secondary sources, the 'search string' field in Endnote is filled with the same details use for its primary source that led to this paper being identified along with words "secondary search". The Field "Name of reference database" is filled in to give information on where search took place, e.g. IEEE *X*plore or ACM.

6.3. Document: Procedure for conducting the search

To ensure that the procedure is reliable and replicable, I and a fellow researcher used the prescriptive process in two separate pilot studies. The outcome of successive trials resulted in the following procedural document which we will use for all our primary searches.

6.4. Document: Procedure for conducting the search

To ensure that the procedure is reliable and replicable, four researchers used the prescriptive process in three separate pilot studies. The outcome of successive trials resulted in the following procedural document which we will use for all our primary searches.

Data

Each one of us performs the systematic review will be given the same Data:

Search Data: URL links to all **databases** on our list (guidelines.doc)

List of **Search terms** tailored to source ("Search-Terms-Database-Name.doc")

Reference Data: Our Research Questions (guidelines.doc)

Exclusion Criteria (guidelines.doc)

Inclusion Criteria (guidelines.doc)

Quality Criteria (guidelines.doc)

The Systematic Procedure (detailed in this section "Checklist of activities")

Output Data:

Results and Quality Assessment Form.doc

For practical purposes, the two output forms ('Quality Assessment' and 'Results' are combined into one document).

Checklist of activities – The Procedure:

INITIAL SET UP

1. Create an electronic 'Patient Privacy Systematic Review' folder, within this folder:

    1.1 Create a 'Patient-Privacy-Generic-documents' folder, in this folder:

        1.1.1 Save 'Search-Terms-Database-Name.doc'

        1.1.2 Save 'Results and Quality Assessment and Form.doc' '

        1.1.3 Save 'Data Synthesis Forms.doc'

    1.2 create a 'Patient Privacy Endnote Library' folder

    1.3 create a 'Patient Privacy Papers' folder

    1.4 create a 'Patient Privacy Secondary Sources' folder

    1.5 create a 'Patient Privacy quality assessments and accepted paper forms' folder

2. Create a 'Papers Assessment' folder with in this folder

    2.1 Create "First Filtered Papers" folder;

    2.2 Create "Pending Decision" folder;

    2.3 Create "Arbitration Papers" folder;

    2.4 Create "Accepted Papers" folder.

3. Open Electronic Reference Manager "Endnote 9".

4. In Endnote, create four Libraries; save them in 'Patient Privacy Endnote Library' folder:

    4.1. All References.enl – a permanent store of ALL downloaded papers;

    4.2. WIP Papers.enl – a temporary store for Work in Progress papers;

4.3. Accepted Papers.enl – a subset of 'All References.enl' library papers.

4.4. Multiple Publications.enl – a subset of 'Accepted Papers.enl' library Papers.

REPEAT FOR EACH INDEX DATABASE IN LIST

5. Go to specified source in lookup table, e.g. ACM, IEEE *E*xplorer, etc.

REPEAT FOR EACH RESEARCH QUESTION (RQ) IN LIST

6. Select one RQ Search String from "Search-Terms-Databse-Name.doc"

7. Copy and Paste string into the search box of the indexing database.

8. Download papers into Endnote'All References.enl' library

REPEAT FOR EACH PAPER IN LIST

9. Open paper in your Endnote 'All References.enl' library and manually fill in all compulsory fields **not** automatically completed in the download, e.g. PaperID, name of researcher, research string reference, etc.

10. Check paper and type in answers to each exclusion criterion (given in separate fields) – there are three possible outcomes:

10.1 If all exclusion **criteria are NOT met** – paper is progressed

to inclusion criteria check (11),

- In "All References.enl" go to Decision field and type "paper not excluded".
- Proceed to inclusion criteria check (no.11)

Else –

10.2 If exclusion criteria **undecided**,

- go to Endnote Decision field
- list why paper cannot proceed to inclusion criteria check
- give date and name of reviewer responsible

- *copy* reference to Endnote Library “WIP Papers” (where paper remains until decision is made: see arbitration 4.1.7.2)
- If decision is to reject – Delete paper from WIP library

  and follow instructions for 10.3.
- If decision is to include, Delete paper from WIP library

  and open paper reference in “All References.enl” **and**

  follow instructions given in 10.1.

Else –

10.3 If **exclusion criteria are met – paper is rejected**

- Go to Decision field and put “reject”, date, reviewer responsible, and reason for fail.
- Go to next paper (work from step 9), If at end of list of papers
  - Go to next Research Question (step 6 onwards), if all research questions have been dealt with
    - Go to next index source (step 5 onwards), if all sources have been searched
      - END input

11. Open reference in Endnote ‘All References.enl’. Check and type answers to each **inclusion criterion** (given in separate fields) – there are three possible outcomes:

11.1 If all inclusion **criteria met,**

• In “All References.enl” go to Decision field and type

“inclusion criteria met”

• *Copy* reference to Endnote Library ‘Inclusion Criteria Met’

• download full electronic paper and save in “Patient Privacy Paper”

folder

- store soft copy in the "accepted papers" folder

Else –

11.2 If inclusion **criteria undecided,**

- Go to Endnote Decision field List why paper cannot proceed to quality check give date and name of reviewer responsible
- *copy* reference to Endnote Library "WIP Papers" (where paper remains until decision is made: see arbitration 4.1.7.2)
- If decision is to reject – Delete paper from WIP Endnote library and follow instructions for 11.3.
- If decision is to include, Delete paper from WIP Endnote library **and** open paper reference in "All References.enl" **and** follow instructions given in 11.1.

Else –

11.3 If **inclusion criteria fails**

- Go to Decision field and put "Inclusion criteria failed", date, reviewer responsible, and reason for fail.
- Go to next paper (work from step 9), If at end of list of papers
  - Go to next Research Question (step 6 onwards) if all research questions have been dealt with
    - Go to next index source, if all sources have been searched,
      - END input

12. Open reference in Accepted Papers.enl Library, go to Quality Assessment form (MS Word Document) embedded in Quality Assessment and Results field.

- Fill in all Quality Assessment fields.

12.1 If quality criteria form can be completed,

- Go to Endnote Quality Criteria field and enter "score".

- Complete the Results section of Quality Assessment and Results.doc"
    - If there are any secondary sources referenced in this form, save copy in Secondary Source folder
- save form as "(PaperID) Results Quality Assessment and.doc
- store soft copy of form (to store with full paper in accepted

papers folder)

- insert completed electronic form as an embedded 'object' into Endnote 'Quality assessment and results form' field.
- Go to Endnote Decision field and report "paper accepted"
- save soft copy of paper in "Accepted Papers" folder together with quality assessment and results.doc form

Else –

12.2 If quality criteria is undecided

- Go to Endnote Quality Criteria (score) field and enter "undecided"
- Go to Endnote "Decision field" and list why paper cannot proceed to the next stage
- Place soft copy of paper in "Papers Pending Decision" folder. Write on paper why paper quality is undecided.

Alternatively,

- If paper needs to go to arbitration - move soft copy of paper into "Arbitration Papers" folder. Write on paper why it needs to go to arbitration. (Paper remains here until decision is made: see arbitration section 7.6.1.)

GO TO NEXT PAPER IN LIST – (go back to 9)

End when all Papers in list have been processed

GO TO NEXT RESEARCH QUESTION – (go back to 6)

End when all Research Questions have been processed

GO TO NEXT INDEX SOURCE – (go back to 5)

End primary search when all agreed Index sources have been searched

6.5. Document: Secondary Source

GO to Secondary Source folder. Do individual searches on new references, authors, journals etc. Follow Procedure for stages 7 – 12.

When all primary AND secondary searches have been completed, run search for multiple studies (see section 7.6.2.). If multiple studies are found select only one study for the review (the most recent/most detailed). After searches for multiple studies have been performed and key paper is identified and selected the data extraction procedure ends.

6.6. Document: Guidelines

The generic procedure outlined above, requires "Guidelines" (see Appendix C) in order to implement the process. It includes rules on how to fill in Endnote fields and forms, our exclusion, inclusion and quality criteria.

6.6.1. Arbitration

Papers that may go to arbitration fall into the following categories:

(a) Papers that are in Pending Decision and Arbitration folders

(b) Papers that are in the electronic WIP library (stored there prior to saving a hard copy or rejecting paper – reason for not proceeding is recorded in Endnote 'Decision' field)

(c) Papers that have been not been accepted by all researchers (identified through comparison of 'Accepted Papers.enl' Endnote libraries).

Stage 1: Internal Arbitration: Researchers involved in the data extraction will try to reach an agreement on all papers (whether to include or exclude).If there is still no agreement, the papers go to stage 2, external arbitration.

Stage 2: External Arbitration: If the first internal arbitration fails to reach an agreement PDF(s) of arbitration paper(s) are sent to other researchers who, based on knowledge of our exclusion criteria, inclusion criteria and quality criteria, will make a final decision – whether to accept or reject the paper.

6.6.2. Multiple Publications

Considering all papers in the 'Accepted Papers.enl', searches are made for articles that report the same study. This is done by grouping papers by author (and coauthors). Duplicate work may not be referenced by the author directly therefore papers grouped by author need to be carefully read to uncover possible duplication. Where duplication is found we include only one paper in our review (that we consider to be the best quality – e.g. the most thorough and ideally most up-to-date). Duplicate papers are removed from 'Accepted Papers.enl' and placed the duplicate papers repository in Endnote Library "Multiple Publications.enl". In this way we avoid giving one finding too much prominence. The one remaining paper in 'Accepted Papers.enl' has its 'repeated study' field filled in with "YES"; # of duplicate papers; "stored in "Multiple Publications.enl"

7. Synthesis of Data Extraction

Data synthesis forms will bring together all the findings reported in our Accepted Papers/Follow-up forms (see). The synthesis comprises qualitative lists of findings that will provide broad answers to our research questions. In order to perform sensitivity analysis we categorize the quality, year, study focus, solution type and context.

Following are two data synthesis forms:

Data Synthesis Form 1: Research Question 1 # of papers accepted (completed at end):

	RQ1? Which solutions of patient data privacy and security have been proposed for software system development?						
Paper ID	Quality Score	Year	Research Type	Solution Type	Context	Privacy principles	Stage of software development
Paper ID	Quality Score	Year	Research Type	Solution Type	Context	Privacy principles	Stage of software development
Etc.							

Table 19 Synthesis Table for RQ1

Context	**Stages**	**Solution Type**	**Proposed Solutions**		**Privacy Principles**	**# of papers**
Distributed network	Requirements	Solution	1. Framework 2. Methodology 3. Process	2 1 3	1. Openness & Transparency 2. Accountability 3. Security control	6
		Validation	1. Scheme 2. Framework 3. Model	1 2 1	1. Individual participation 2. Purpose specification	4
		Evaluation	---------------		---------------	------

	Implementation	Solution				
		Validation	1. Technique 2. Method	1 2	1. Individual Participation 2. Purpose Specification 3. Remedies	3
		Evaluation	1. Scheme 2. Technique 3. Model	1 1 2	1. User Consent 2. Collection Limitation 3. Security Control	4
			——————		——————	————
	Design	Solution	1. Framework 2. Methodology 3. Model 4. Method			
		Validation	1. Process 2. Framework 3. Model			
		Evaluation				
	Testing	Solution				
		Validation				
		Evaluation				
Etc						

Table 20 Data Synthesis Table for RQ2

Where, proposed solutions refer to any technique, process, methodology, method, model etc stages will be any software development stage and type of paper shows which type research had been conducted.

Following are the description of search types adopted from[125]:-

Category	Description
Validation Research	"Techniques investigated are novel and have not yet been implemented in practice. Techniques used are for example experiments, i.e., work done in the lab ".
Evaluation Research	"Techniques are implemented in practice and an evaluation of the technique is conducted. That means, it is shown how the technique is implemented in practice (solution implementation) and what are the consequences of the implementation in terms of benefits and drawbacks (implementation evaluation). This also includes identifying problems in industry".
Solution Proposal	"A solution for a problem is proposed, the solution can be either novel or a significant extension of an existing technique. The potential benefits and the applicability of the solution are shown by a small example or a good line of argumentation ".
Philosophical Papers	"These papers sketch a new way of looking at existing things by structuring the field in form of a taxonomy or conceptual framework ".
Opinion Papers	"These papers express the personal opinion of somebody whether a certain technique is good or bad, or how things should been done. They do not rely on related work and research methodologies ".
Experience Papers	"Experience papers explain on what and how something has been done in practice. It has to be the personal experience of the author".

Table 21 Research Facet Categorization

8. Validation of Protocol

Initially the work has been started by Asma Naveed and Umrah Naeem for their semester report. I have taken their idea for my final research thesis of MS. I have review the protocol and previous pilot study but found lots of conflicts and problems especially in narrowing down the focus in research questions. I have tailored the protocol and validated it from experts. I have performed systematic mapping study which really helped me in narrow down my focus for my research questions. This study explored what type of contributions has been made in this area which I have addressed in RQ1 to cover maximum work done on patient data privacy and security for software system development. After performing Systematic mapping study, I have developed protocol from the scratch especially focusing on the suggestions given by Barbara Kitchenham. Protocol was tailored three times on feedback of my supervisor Dr. Saad Zafar and finally was validated by students.

5.1. Process

First phase (pilot study)

Two separate search strings are run which give lots of paper but time doesn't allow refining all of them only 10 are selected to validate the protocol the purpose of it not to show completeness but consistency in results and validate the protocol in auditable and repeatable manner.

Consistency in results

Papers found in database search: 2 sets of research terms are used in IEEE explorer.

TITLE-ABSTR-KEY(("Health Records" OR " Patient Information" OR "Medical Records" OR " Electronic Healthcare Records" OR " Electronic Health Records" OR " Electronic Medical Records" OR " Protected Health Information" OR " Healthcare Records" OR "Health Record System")) and TITLE-ABSTR-KEY((Privacy OR Security))

TITLE-ABSTR-KEY (("Healthcare Records" OR "Health Record System")) and TITLE-ABSTR-KEY((Privacy OR Security))

Result: It is important that this category is the same for both of us. There is a close to 100% agreement on the papers downloaded from the database using the given search terms.

All_references Library: All papers downloaded from database in this library, not pass full inclusion and exclusion criteria these all will be not included in our final review.

Result:

There are difference in numbers of papers downloaded well all of these will be not included in final review .Differences pointed out in 'paper _accepted' library.

Work in Progress (WIP) library: This library is a temporary.

Store for papers that need more information (e.g. full papers) before a decision can be made.

Result: There are 2 papers initially stored in this library when discussion made both are deleted.

Papers that meet the inclusion criteria: Papers that meet inclusion criteria (answers a RQ, reliable source)

Result: There was only one paper difference.

Second phase (Expert opinion)

In second phase of protocol validation the protocol is send to experts in the field of Systematic Literature Review (Barbara Kitchnham and Dr. Mehmood Niazi).

6. Schedule of Activities

Table 27 shows schedule of activities for developing prortocol

Activity	Date	People Involved	Completion Date	Comments
Pre-SPilot(1)	15-03-2010	Asma, Umrah,Isma	05-04-2010	Completed
Pilot	06-04-2010	Asma, Umrah,Isma	30-04-2010	Completed
Protocol developed v1	10-05-2010	Isma	25-06-2010	Completed
Protocol circulated for comments	26-06-2010	Barbara Kitchenham, Dr. Mehmood Niazi Dr. Naveed Ikram Dr. Iftikhar Niazi	27-06-2010	Completed
Systematic mapping	01-07-2010	Isma	30-10-2010	Completed
Protocol V2	01-11-2010	Isma	30-11-2010	Completed
Comments received	25-12-2010	Barbara Kitchenham, Dr.Mehmood Niazi Dr.Naveed Ikram Dr.Iftikhar Niazi	05-01-2011	Completed
Protocol v3	20-01-2011	Isma	02-02-2011	Feedback from supervisor
Protocol v4	28-02-2010	Isma	03-03-2011	Completed

Table 22 Schedule of Activities

Appendix A: Example of Guidelines for implementing the Review Process

Research Questions

RQ.1 Which solutions of patient data privacy and security have been proposed for software system development?

Population: Empirical Literature for solutions of patient data privacy and security.

Outcome: Solutions for patient data privacy and security propose for software system development.

Intervention: Solutions to answer privacy principles – Stages of software system development

Comparison: No

RQ.2. Can we categorize these solutions using the Markle Foundation's Common Framework?

Population: List of empirical solutions on patient data privacy and security for software development.

Outcome: Solutions for patient data privacy and security propose for software system development to answer privacy policy principles.

Intervention: Solutions to answer privacy principles – Stages of software system development

Comparison: on the basis of privacy policy principles.

RQ1 (copy and paste the search terms)

((Health Records" OR" Patient Information" OR "Medical Records" OR" Electronic Healthcare Records" OR" Electronic Health Records" OR" Electronic Medical Records" OR" Patient Data" OR" Medical Data" OR" Protected Health Information") AND (Privacy OR Security))

Please note down the search terms tailored for each database here:-

Lookup table1

<table>
<tr><th>Date</th><th>Search String
TI=within title| AB=within Abstract| All= within ALL</th><th>Search Identifier</th><th>Comment</th></tr>
<tr><td rowspan="8"></td><td rowspan="8"></td><td></td><td>RQ1-search</td></tr>
<tr><td colspan="2">4 separate searches</td></tr>
<tr><td>IEEE Explorer</td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
</table>

Lookup table 2

<table>
<tr><th>Date</th><th>Search String
TI=within title|AB=within Abstract| All= within ALL ,*=truncuation</th><th>Search Identifier</th><th>Comment</th></tr>
<tr><td rowspan="8"></td><td rowspan="8"></td><td></td><td>RQ1-search</td></tr>
<tr><td colspan="2">4 separate searches</td></tr>
<tr><td>ACM</td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
</table>

Lookup table 3

<table>
<tr><th>Date</th><th>Search String
TI=within title|AB=within Abstract| All= within ALL ,*=truncuation</th><th>Search Identifier</th><th>Comment</th></tr>
<tr><td rowspan="7"></td><td rowspan="7"></td><td></td><td>RQ1-search</td></tr>
<tr><td colspan="2">4 separate searches</td></tr>
<tr><td>SpringerLink</td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
</table>

Lookup

<table>
<tr><th>Date</th><th>Search String
TI=within title|AB=within Abstract| All= within ALL ,*=truncuation</th><th>Search Identifier</th><th>Comment</th></tr>
<tr><td rowspan="7"></td><td rowspan="7"></td><td></td><td>RQ1-search</td></tr>
<tr><td colspan="2">4 separate searches</td></tr>
<tr><td>Science Direct</td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
</table>

2. Endnote Fields:

We have used one Reference type for all sources (Journals/Conferences/Web material)

Following are some instructions by which we can modify our Endnote Library:-

7.1.1.1.1 To rename a field:

1. From the *Edit* menu, choose *Preferences*, select the *Reference Type* option in the list of preferences, and click *Modify Reference Types* to open the Reference Types preference.
2. Use the drop-down list at the top to find the reference type that you want to change.
3. Within the column for that reference type, find the field name that you want to change, click on it, and type a new name for the field to replace the current name.
4. (Optional) If you want to change the field for all reference types, click the *Apply to All Ref Types* button.
5. Click *OK* to return to the main Preferences window for Reference Types.
6. Click *OK Save* to save your changes.

Styles, filters, and connection files update automatically to use the new name.

7.1.1.1.2 To add a field to a reference type:

1. From the *Edit* menu, choose *Preferences*, select the *Reference Type* option in the list of preferences, and click *Modify Reference Types* to open the Reference Types preference.
2. Use the drop-down list at the top to find the reference type that you want to change.
3. Look at the field names listed in the Generic column and find the one with the most similar meaning to the field that you want to add. Make sure that the corresponding cell is blank for the reference type that you are modifying. If it is not blank, then you should use another field.
4. Click in the blank cell and type the name for the new field.
5. (Optional) If you want to add the new field to all reference types, click the *Apply to All Ref Types* button.

6. Click *OK* to return to the main Preferences window for Reference Types.
7. Click *OK Save* to save your changes.

7.1.1.1.3 To delete a field from a reference type:

1. From the *Edit* menu, choose *Preferences*, select the *Reference Type* option in the list of preferences, and click *Modify Reference Types* to open the Reference Types preference.
2. Use the drop-down list at the top to find the reference type that you want to change.
3. Find the name of the field you want to delete and select it.
4. Press the Delete or Backspace key to clear that field name.
5. (Optional) If you want to delete the field from all reference types, click the *Apply to All Ref Types* button.
6. Click *OK* to return to the main Preferences window for Reference Types.
7. Click *OK Save* to save your changes.

The deleted field no longer appears in any references using that reference type. However, if there was any information in the deleted field, it still appears in the reference, but the field is displayed with its Generic name. For example, suppose you remove the Editor field from the Book reference type. Thereafter, when you add new book references to your library, there will be no available field for entering an editor. However, if you edit an old book reference, one in which you had entered an editor's name, the name will be displayed in the field titled Secondary Author. It is the same Editor field that was used originally, however it is now displayed with its Generic name.

Data in a field is not deleted by deleting a field from a reference type format. To remove all text from a field, use the "Clear Field" option in the *Change Field.*

8 Appendix 2: Publication

proposed for software system development? (2) Can we categorize these solutions using the Markle Foundation's Common Framework?

In Section II, we have described our systematic mapping process; in Section III, we provide explicit answers to our research questions; the discussion of the results is provided in Section IV; conclusion and the future work are given in the last section.

II. THE SYSTEMATIC MAPPING PROCESS

For our mapping study, we following the guidelines provided in [7, 8]. Accordingly, our mapping study was conducted in three stages. In Stage 1, we define the scope, the search strategy and the selection criteria. In the second stage primary studies were selected applying the search strategy and the selection criteria. Lastly, in Stage 3, the selected studies are classified into the different categories.

A. Stage 1: Defining Scope, search strategy and selection criteria

We define the scope of the study as follows. The *population* of the study is selected as the set of articles addressing patient data privacy and security. As *intervention*, we selected any patient data privacy and security solution proposed for any of the software development cycle (e.g., requirements engineering, design, testing, etc.). The *outcome* of our study is a mapping of selected solutions to the patient data privacy principles found in [9]. Our search string for conducting the research was:

Patient AND *Data* AND (*Privacy* OR *Security*)

The research sources selected for our study were IEEE Digital Library, ACM Digital Library, Science Direct and Springerlink. To select relevant studies, we used the following inclusion and exclusion criteria.

Inclusion Criteria: A study contribution related to any stage of the software system development lifecycle. The study should also discuss *at least* one or more than one principles of patient data policy. For this purpose we read abstract, conclusion, introduction, or the full paper (if required).

Exclusion Criteria: Any study not related to the domain of software engineering, patient data privacy or security is not selected. The studies related to patient data privacy and security for images, sensor network and wireless transmission are also not included.

B. Stage 2. Selecting primary studies

In the first iteration, the search string was used at each resource. All references along with their abstracts were downloaded in Endnote [11] reference library. At this stage, we downloaded 4,670 references. In the second iteration, abstract of all reference were read and relevant studies which explicitly addressed the patient data privacy or security with contribution towards software system development were selected and placed in another library of selected papers. In this iteration, 120 studies were selected. We selected 93 papers from IEEE, 6 papers from ACM, 17 papers from Science Direct and 4 papers from Springerlink. In the third iteration, full texts of these 120 studies were downloaded. We read all the articles one by one and applied the inclusion and exclusion criteria and finally selected 51 studies in our third iterative phase. We placed our 12 doubtful studies in the pending folder. In the fourth iteration, we discussed these doubtful studies and decided to accept 7 studies and to reject 5 studies. The breakdown of the results from each of the source is presented in Table 1, whereas Table 2 shows the distribution of our four iterative phases and the number of studies which were retained in each phase. In Table 3, we summarize the most relevant publication channels.

TABLE 1. NO. OF STUDIES AT EACH RESOURCE

Resource	No. of studies	No. of selected studies	Percentage
IEEE	4,540	44	0.96%
ACM	74	6	8.1%
Science Direct	40	8	20%
Springerlink	16	0	0%
Total	4,670	58	1.2%

TABLE 2. NO. OF STUDIES AT ITERATIONS

1st iteration	2nd iteration	3rd iteration	4th iteration
4,670	120	51	58

TABLE 3. MOST RELEVENT PUBLICATION CHANNELS

Acronym	Type of publication	Percent
International Journal of medical informatics	Journal	13.7%
Information Technology in Biomedicine	Journal	6.8%
CCSW	Workshop	5%
ICBECS	Conference	3.4%

The IEEE Digital Library had yielded the most number of papers (4,670), followed by ACM (74), Science Direct (40), and Springerlink (16). It is noteworthy that the most relevant studies were found in Science Direct (20%) and the least were found in Springerlink (0%). ACM had 8.1% and IEEE Digital Library had 0.96% relevant studies, respectively. Most of the relevant studies were found in International Journal of Medical Informatics (13.7%). This was followed by Information Technology in Biomedicine (6.8%). The rest of the relevant studies were found in two conferences: Workshop on cloud computing security (CCSW) (5%) and International Conference on Biomedical Engineering and Computer Science (ICBECS) (3.4%).

As part of our inclusion criteria, we included studies from the year 2000 to 2011. For the year 2000 we did not find any relevant study. However, from the years 2001 to 2008 the number of relevant studies increased steadily with a sharp increase in the year 2008 (frequency=17). The only exception to the trend is the year 2009 where the total number was reduced to only 4. In 2010 the number was again increased to 10 studies showing a positive trend. Only one study was found to be relevant in the first quarter of

2011. This trend of number relevant studies per year is given in Table 4.

TABLE 4. PERCENTAGE OF STUDIES AT EACH YEAR

Years	Relevant Studies	Selected Studies	Percentage
2000	2	0	0%
2001	5	2	3.4%
2002	6	1	1.7%
2003	8	3	5.1%
2004	8	3	5.1%
2005	10	3	5.1%
2006	10	4	6.8%
2007	23	8	13.7%
2008	25	17	29.3%
2009	30	4	6.8%
2010	23	10	17.2%
2011	22	1	1.7%
Total	**172**	**58**	

C. *Stage 3. Classifying selected Studies*

In the next stage, we divided our studies according into three categories. In the first category, we classified the studies according to the research approach used in the selected primary studies. We divide the research approaches according to the classification proposed by Weiringa et al. [10]. The *validation research* is used for those novel techniques that have not been implemented and are validated through experiments in a lab-like environment. The *evaluation research* is used to evaluate the techniques that have been implemented in practice. This research type explores how well the technique has been implemented. In the *solution proposal* either a novel solution is proposed or an existing solution is extended significantly. The *philosophical papers* propose either a conceptual framework to structure concepts into a new taxonomy. On the other hand *opinion papers* express personal opinion of the authors about a technique and the *experience papers* explain the experience of the authors of how a technique has been implemented in practice.

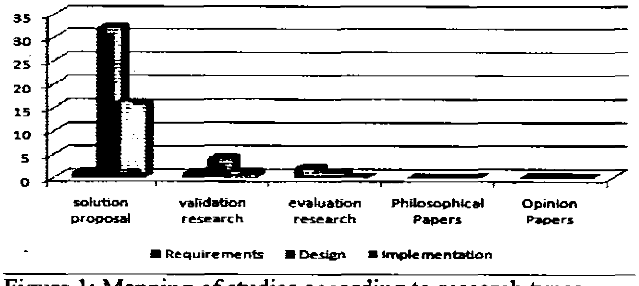

Figure 1: Mapping of studies according to research types

TABLE 5. RESEARCH TYPE AND SOFTWARE DEV.PHASE FACETS

Context	Solution	Validation	Evaluation	Total
Req.	1	1	2	4
Design	32	4	1	37
Imp.	16	1	0	17
Ver.	0	0	0	0
Maint.	0	0	0	0
Total	**49**	**6**	**3**	**58**

Table 5 shows the distribution of research type facet of the selected studies. An overwhelming majority of research approaches in the selected primary studies proposed a new solution (*f*=49). The next approach used the most was validation research (*f*=6) followed by evaluation research (*f*=3). However, we did not find any study that could be classified into any of the other research type categories. The results of this classification are summarized in Figure 1.

We also classified the studies on the basis of different stages of software development. Specifically, we grouped the software development stages into: *requirements, design, implementation, verification,* and *maintenance.* The breakdown of the classification of the selected studies is given in Table 5. The majority of selected primary studies addressed the Design phase of the software development (*f*=37), followed by the Implementation phase (*f=17*), while some of the studies were classified under the Requirements phase (*f*=4). We did not find any study related to software Verification and Maintenance phases.

Our next categorization was based on the Markle Foundation's privacy principles [9]. The first principle of (1) *Openness and Transparency* mandates that there should be an overall policy of openness regarding personal data. The individuals should be aware of the nature stored data, its location and its access control policy. The (2) *Purpose Specification and Minimization* principle requires that the data collection purpose should be defined at the time of collection and its use should be limited to the intended purpose. Under the (3) *Collection Limitation* principle the personal health information must only be collected lawfully and with the knowledge and consent of the concerned individual. The (4) *Use Limitation* principle states that personal data must not be disclosed, made available or used in any manner other than the specified purposes. The (5) *Individual Participation and Control* principle requires that individuals have the right of access and control over their stored personal information. The (6) *Data Integrity and Quality* states that only the relevant data is stored and that the data is always accurate, complete, and current. The (7) *Security Safeguards and Controls* requires there should be reasonable security safeguards against the risks of loss of data or unauthorized access. The accountability of entities responsible for keeping and maintaining the personal data according to stated principles is covered under the (8) *Accountability and Oversight* principle. Lastly, the (9) *Remedies* principle states that there are adequate legal and financial remedies to address any security breaches or privacy violations.

Table 6 shows the distribution of studies according to the aforementioned privacy principles. As reflected in the data shown in the table, we found many single studies that address multiple privacy principles. The most coverage was given to the Use Limitation principle (*f*=38). This was followed equally by the Individual Participation and Control, and Security Safeguard and Control principles (*f*=24). After them the most covered principle was Data Integrity and Quality principle (*f*=16), followed by the Purpose Specification Principle (*f*=14). The next principle covered the most was the Accountability and Oversight principle (*f*=13), whereas, the Remedies and Collection Limitation had the least coverage with a frequency of 3 and 1, respectively.

TABLE 6. CLASSIFICATION OF STUDIES ACCORDING TO PRIVACY PRINCIPLES

Principle	Req.	Design	Impl.	Total
Openness	2	5	1	8
Purpose Specification	1	9	4	14
Collection Limitation	1	0	0	1
Use Limitation	1	23	14	38
Individual Participation and Control	1	17	6	24
Data Integrity and Quality	1	13	2	16
Security Safeguards and Control	2	17	5	24
Accountability and Oversight	2	9	2	13
Remedies	0	2	1	3

III. RESEARCH QUESTIONS

Based on the above data, we now answer our two research questions.

RQ1: Which solutions of patient data privacy and security have been proposed for software system development?

. In our mapping study we found 58 relevant primary studies. Out of these studies 63% of the studies were related to the Software Design. While 27% of the studies contributed towards Software Implementation and only 6% aimed at Software Requirements. Therefore, we can conclude that the most research is being conducted on how to effectively design software systems related to the requirements of patient data privacy and security. Similarly, there is also significant focus in the research community on how to effectively implement the patient data privacy and security requirements. Surprisingly, much less studies are focused on requirements analysis and specification phase of software development (see Figure 2).

RQ2: Can we categorize these solutions using the Markle Foundation's Privacy Principles [9]?

The mapping of selected studies against the Markle Foundation's Privacy Principles is given in Figure 3. As discussed earlier, a single study was often mapped against multiple principles. But we found the solutions in the studies mapped reasonably well against the privacy principle. It is important to note that the Use Limitation was covered in 41.4% of the studies, followed by Individual Participation and Security Control principles with 41.4% studies. The other two principles covered in the selected studies were Data Integrity and Quality, and Purpose Specification with 27.6% and Purpose Specification 24.1%, respectively. The coverage of rest of the principles was not very significant.

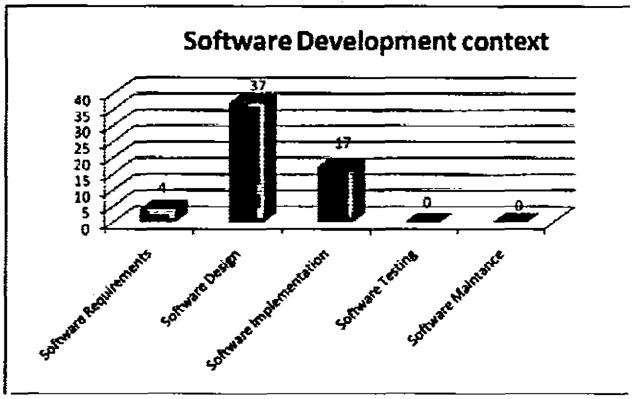

Figure 2: Mapping of studies according to software development context

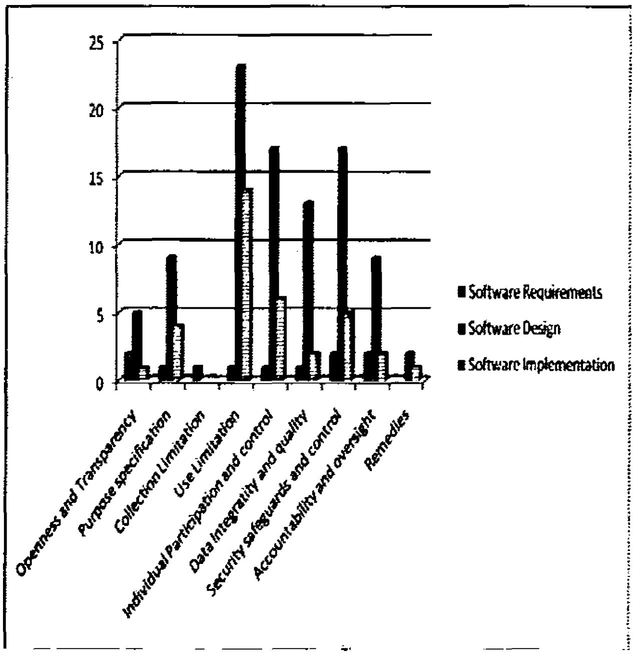

Figure 3: Mapping of studies against privacy principles

IV. DISCUSSION

The amount of personal information stored and exchanged by the health information systems is increasing by the day. With the increase in the volume of data the concern about the patient data privacy and security is also increasing. The data stored about the patient include sensitive information like history of diseases and treatments, history of drugs used, sexual orientation and practices, results of sexually transmitted diseases, etc. As a result, a number of rules, regulations and best practices have been proposed to ensure that the stored data does not violate individual's privacy and that the data is never use inappropriately. Consequently, there has been a steady increase in research community to ensure that the software systems deployed must effectively integrate all the requirements related to patient data privacy and security.

The motivation behind our study was to investigate the feasibility for conducting a complete Systematic Literature Review. Here we cover the breadth of patient data privacy and security presented in the literature. The subsequent SLR studies can investigate the depth based on the results presented in our work.

The steady increase in the related primary studies from the year 2001 to 2010, with a few possible exceptions, indicates a growing interest in this significantly important research area (see Table 4). Similarly, the need of implementation of patient data and security requirements is reflected from the fact that most of the selected studies are concerned about the Design and Implementation of the privacy related requirements and less attention is paid to critically important phases of Requirements Analysis and Specification, Verification and Maintenance. This notion is further reinforced by the fact that the most common research approach used in the primary studies is Solution Proposal, with much less studies on validation and evaluation research. Likewise, we did not find any study based on experience reports, philosophical papers, or opinion papers.

Perhaps, not surprisingly the most importance is given to the Use Limitation, Individual Participation and Security Control principles. However, less coverage is given to the rest of the privacy principles, without which any software system cannot effectively implement a complete set of patient data privacy and security requirements.

We identify the following two limitations of our study: (1) some studies may have been missed due to the diverse use of the terms used in the search string; and (2) studies published in English language were selected in the search.

V. CONCLUSION AND FUTURE WORK

In this study, we have presented initial findings on solutions available for patient data privacy and security to develop software system. On this topic, we found 58 papers published in the years from 2000 to the first quarter of 2011. We have mapped these solutions against principles of privacy policy to cover all aspects of patient data privacy and security. A large number of studies focused on Software Design as compared to Software Implementation and Software Requirements while, no study found on testing and maintenance. The Use Limitation principle along with Individual Participation and Control, and Security Safeguard and Control had most coverage in the selected studies. Our future work includes performing in-depth Systematic Literature Review on various aspects of Patient Data Privacy and Security identified in this study.

REFERENCES

[1] U S. Congress, Office of Technology Assessment," Protecting Privacy in Computerized Medical Information" OTA-TCT 576. Washington, DC, US Government Printing Office, Sept 1993.

[2] A. Appari and M.E. Johnson., "Information Security and Privacy in Healthcare: Current State of Research."*International Journal Internet and Enterprise Management*, vol. 6, pp. 279-314, Oct. 2010.

[3] L.Gostin., "Health Care Information and Protection Privacy : Ethical and Legal Considerations" in ETATS-UNIS, 1997, pp. 683-690.

[4] C. H. Liu, Y. F. Chung, T. S Chen, and S. D Wang, "The Enhancement of Security in Healthcare Information Systems." *International Journal of Medical System*",pp. 1-16, Nov. 2010.

[5] M.Eichelberg, T. Aden, J. Riesmeier, A. Dogac, and G. B. Lalecil., "A Survey and Analysis of Electronic Healthcare Records Standards. "*Journal ACM Computing Surveys*, vol.37, pp. 277-315, Dec. 2005.

[6] B. Kitchenham and S.Charters., "Guidelines for performing systematic literature reviews in software engineering", Technical Report, EBSE-2007-01, Keele University, 2007.

[7] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson., "Systematic mapping studies in software engineering.", in 12th International Conference on Evaluation and Assessment in Software Engineering (EASE), pp. 71-80 , June. 2008.

[8] W. Afzal, R. Torkar, and R. Feldt., "A systematic mapping study on non-functional search-based software testing", in 20th International Conference on Software Engineering and Knowledge Engineering (SEKE), 2008.

[9] Markle Foundation, Connecting for Health Common Framework. January 10, 2011. <www.connectingforhealth.org>

[10] R.Wieringa, N.Maiden, N.Mead, and C.Rolland, "Requirements engineering paper classification and evaluation criteria: a proposal and a discussion", *Journal Requirements Engineering* . vol. 11, pp. 102–107, Dec. 2005.

[11] T.Reuters,"EndNote-Your smater refrence assistant"Internet. June 5, 2010. <http://www.endnote.com/>

[62] O'Kefee C, Greenfield P, and Goodchild A, "A Decentralized Control Approach to Electronic Consent and Health Information Access Control," *Journal of Research and Practice in Information Technology* vol. 37, pp. 161-178, 2005.

[63] Bhasker R, Laxman S, Smith A, and et al, "discovering frequent patterns in sensitive data," *KDD'10 16th Int Cong,* pp. 503-512, 2010.

[64] Boehm O, Caumanns J, Franke M, and et al, " Federated authentication and authorization: A case study," *Enterprise Distributed Object Computing Conf,* pp. 356-362, 2008.

[65] Bolchini C and Schrciber F A, "Smart card embedded information system a methodology for privacy oriented architectural design," *Data & Knowledge Engineering,* vol. 41, pp. 159-182, 2002.

[66] Brumen B, Welzer T, Druzove M, and et al, " protecting medical data for analyses," *CBMS Symp Proc,* pp. 102-107, 2002.

[67] Chi H G, Jous L E, and Zhao L, "Implementation of security access control model for inter-organizational healthcare information systems," *APSC Conf,* pp. 692-696, 2008.

[68] Compagn L, Khoury E P, Krausova A, and et al, "How to integrate legal requirement engineering methodology for the development of security and privacy patterns," *Artificial Intelligence and Law,* vol. 17, pp. 1-30, 2008.

[69] Danilatou V and Ioannidis S, "Security and privacy architecture for biomedical cloud computing," *ITAB 10th Conf,* pp. 1-4, 2010.

[70] Demuynck L, Decker D B, and "Privacy-preserving electronic health records," *communication and multimedia security,* vol. 3677, pp. 150-159, 2005.

[71] Dwivedi A, Bali K R, Belsis A M, and et al, "Towards a practical healthcare information security model for healthcare institutions," *Information Technology Application in Bio medicine 4th Int Conf* pp. 114-117, 2003.

[72] Ferreira A, Antunes L, Chadwick D, and et al, "Grounding information security in healthcare," *International Journal of Medical Informatics,* vol. 19, pp. 268-283.

[73] Finance B, Medjdoub S, and Pucheral P, " Privacy of medical records: from law principles to practice," *CBMS Symp Proc,* pp. 220-225, 2005.

[74] Gritzalis S and Lambrinoudakis C, "A security architecture for interconnecting health information system," *International Journal of Medical Informatics,* vol. 73, pp. 305-309, 2004.

[75] Hadzic M and Dillon T, "Use of ontology technology for standardization of medical records and dealing with associated privacy issues," *ICIT Int Conf* pp. 2839-2845, 2006.

[76] Han S, Skinner G, Potdar, and et al, "New framework for authentication and authorization for e-health service systems," *ICIT Int Conf,* pp. 2833-2838, 2006.

[77] Hass S, Wohlgemuth S, Echizen I, and et al, "Aspects of privacy for electronic health records," *International Journal of Medical Informatics,* vol. 80, pp. 26-31, 2011.

[94] Riva A, Mandl D K, On H D, and et al, "The personal internetworked notary and guardian," *International Journal of Medical Informatics* vol. 62, pp. 27-40, 2001.

[95] Song J and Chang M, "SHOES : Secure Healthcare Oriented Environement Service Model," *BioCas conf*, pp. 89-93, 2006.

[96] Xiao L, Vicente J, Saez C, and et al, "A security model and its application to distributed support system for healthcare," *Availability, Reliability and Security 3rd Int Conf* pp. 578-585, 2008.

[97] Boyd D A, Hosner C, Hunscher A D, and et al, "A 'Honest Broker' mechanism to maintain privacy for patient care and academic medical record," *International Journal of Medical Informatics*, vol. 76, pp. 407-411, 2007.

[98] Bhattacharya J , Gupta K S, and Agrawal B, " Protecting privacy of health information through privacy broker," *System Sciences 39th Int Conf Proc*, pp. 1-10, 2006.

[99] Chen Q, Wang Z, and Zhang W, "Security design for electronic medical record sharing system," *ICBECS Int Conf*, pp. 1-4, 2010.

[100] Lien C, Li W C, and Chen A Y, "Integrating security considerations in client server architectures of health information system development," *IMIS 5th Int Conf*, pp. 527-532, 2011.

[101] Massey K A, Oho N P, Hayward, and et al, "Evaluating existing security and privacy requirements for legal compliance," *Requirements Eng*, vol. 15, pp. 119-137, 2010.

[102] Mchumo S and Chi H, "A framework for access control model in enterprise Healthcare via SAML," *ACMSE'10 48th ANN Conf*, pp. 1-2, 2010.

[103] Rohmouni B, Solomonides H, Mont C T, and et al, "Privacy compliance in European health grid domains: An ontology based approach," *CBMS Symp*, pp. 1-8, 2009.

[104] Sun J and Fang Y, " Cross domain data sharing in distributed electronic health record system," *Parallel and distributed electronic health record system*, vol. 21, pp. 754-764, 2010.

[105] Reni G, Molteni M, Arlotti S, and et al, "Chief medical officer actions on information security in an Italian rehabilitation centre," *International Journal of Medical Informatics*, vol. 73, pp. 271-279, 2004.

[106] Sucurovic S, "Implementing security in a distributed web-based EHCR," *International Journal of Medical Informatics* vol. 76, pp. 491-496, 2007.

[107] Weaver A C, Dwyer S J, Snyder A M , and et al, "Federated secure trust networks for distributed healthcare IT services," *INDIN conf Proc* pp. 162-169, 2003.

[108] Ion M, Russello G, and Crispo B, "Enforcing multi-user access policies to encrypted cloud databases," *Distributed System and Networks Symp*, pp. 175-177, 2011.

[109] Suliman R, Sharma D, Wan li M, and et al, "Security architecture for e-health services," *ICACT 10th Int Conf*, pp. 999-1004, 2008.

[110] G. M. Ueckert F , Ataian M, et al, "Empowerment of Patient and Communication with Healthcare Professionals Through an Electronic Health Record," *Int Journal of Medical Informatics* vol. 70, pp. 99-108, 2003.

[111] Malin B, " Secure construction of K-unlinkable patient records from distributed providers," *Artificial Intelligence in Medicine* vol. 48, pp. 29-41, 2010.

[112] Rashid H A and Hagazy A F, " Protect privacy of medical informatics using K-anonymization model," *INFOS 7th Int Conf* pp. 1-10, 2010.

[113] Khatri A, Kabra S, Singh S, and et al, " Architecture for preserving privacy during data mining by hybridization of partitioning of medical data," *AMS Int Conf*, pp. 93-97, 2010.

[114] Loukides G, Divanis G A, and Malin B, " COAT: Constraint-based anonymization of transactions," *Knowledge Inf Syst*, vol. 28, pp. 251-282, 2011.

[115] Wang Q and Jin H, "Quantified risk adaptive access control for patient privacy protection in health information system," *ASIACCS'11 6th Symp*, pp. 406-410, 2011.

[116] Geneiatakis D, Lambrinoudakis C, and Gritalis S, "A hierarchical model for cross domain communication of healthcare units," *NSS 3rd Int Conf*, pp. 123-129, 2009.

[117] C.-H. Liu, Y.-F. Chung, T.-S. Chen, and S.-D. Wang, "The Enhancement of Security in Healthcare Information Systems," *Journal of Medical Systems*, pp. 1-16.

[118] B. A. Kitchenham, "Guidelines for performing Systematic Literature Reviews in Software Engineering," *Software Engineering Group School of Computer Science and Mathematics Keele University*

pp. 1-57, 2007.

[119] P. S. P. S. Lucila Ohno-Machadoa, Staal Vinterboa, "Protecting patient privacy by quantifiable control of disclosures in disseminated databases," *International Jouranl of Medical Informatics*, vol. 73, pp. 599—606., 2004.

[120] M. E. J. Ajit Appari "Information security and privacy in healthcare: Current state of research," *International J. Internet and Enterprise Management*, 2009.

[122] J. Lawrence Gostin, "Health Care Information and the Protection of Personal Privacy: Ethical and Legal Considerations," vol. 127, pp. 683-690 1997.

[123] E. Marco, et al, "A survey and analysis of Electronic Healthcare Record standards.," *ACM Comput. Surv*, vol. 37(4), pp. 277-315, 2005.

[124] N. B. Sarah Beecham, Tracy Hall,Hugh Robinson2 ,Helen Sharp, "Protocol for a Systematic Literature Review of Motivation in Software Engineering,Technical Report No: 453," *School of Computer Science, University of Hertfordshire,College Lane Campus, Hatfield, Hertfordshire*, 2006.

[125] R. Wieringa, Maiden, N. A. M., Mead, N. R. & Rolland, C. , "Requirements engineering paper classification and evaluation criteria: a proposal and a discussion',." *Requir. Eng*, vol. 11(1), pp. 102–107, 2006.